REMARKS

Claims 1-36 are pending. Claims 6, 7, 10-20, and 24-31 have been withdrawn from consideration. Claim 1 is the sole independent claim that is under consideration.

Objections to the Specification

In the Office action mailed February 14, 2008, the title of the application was objected to as not being descriptive. As shown above, the title has been amended to address the Examiner's concerns. Applicant respectfully requests that the objection to the title be withdrawn.

Objections to the Claims

Claims 8 and 23 were objected to as including a typographical informality. As shown above, claims 8 and 23 have been amended to address the Examiner's concerns. Applicant respectfully requests that the objection to the title be withdrawn.

Rejections under 35 U.S.C. § 101

Claim 1 was rejected under 35 U.S.C. § 101 as allegedly directed to non-statutory subject matter. As shown above, claim 1 has been amended to recite that the model is made available to estimate the sensitivity of a biological specimen to a specific drug. Accordingly, applicant respectfully requests that the rejection under 35 U.S.C. § 101 be withdrawn.

Rejections under 35 U.S.C. § 112

Claim 1 was rejected under 35 U.S.C. § 112, second paragraph, as allegedly indefinite. The rejection contends that the scope of the recitation that a model can be constructed by a "partial least squares method type 1 using said sensitivity data [and gene expression data]" is not discernable to those of ordinary skill.

Applicant respectfully disagrees and submits that those of ordinary are able to discern how a model can be constructed by a partial least squares method type 1 "using" sensitivity data and gene expression data. For example, submitted herewith are pages 293–359 from the book entitled "Statistical Procedures in Food Research." The chapter entitled "Partial Least Squares Regression," by M. Martens and H. Martens (hereinafter "Martens"), illustrates conceptual differences between multiple linear regression and partial least squares in FIG. 2 (page 300). In partial least squares type 1, "**Y** is used as a guide in the extraction of latent variables **T** from **X**..." (underlining added).

Since the term "using" is well understood by those of ordinary skill, especially in the context of constructing models, Applicant submits that the scope of claim 1 is discernable to those of ordinary skill and respectfully requests that the rejection of claim 1 under 35 U.S.C. § 112, second paragraph, be withdrawn.

Claims 1, 2, 3, and 5 were rejected under 35 U.S.C. § 112, second paragraph, as allegedly indefinite. The rejections contend that these claims recite "passive steps."

Although applicant disagrees, as shown above, claims 1, 2, 3, and 5 have been amended to address the Examiner's concerns. Accordingly, applicant respectfully requests that the rejections under 35 U.S.C. § 112, second paragraph, be withdrawn.

Claims 2, 3, and 4 were rejected under 35 U.S.C. § 112, second paragraph, as allegedly indefinite on various ground. As shown above, claims 2, 3, and 4 have been amended to address the Examiner's concerns. Applicant respectfully requests that the rejections of claims 2, 3, and 4 under 35 U.S.C. § 112, second paragraph, be withdrawn.

Rejections under 35 U.S.C. § 102

Claim 1 was rejected under 35 U.S.C. § 102(e) as anticipated by U.S. Patent Publication No. 2002/0111742 to Rocke et al. (hereinafter "Rocke").

As amended, claim 1 relates to a method for constructing a model that estimates sensitivity to a drug based on expression levels of genes. The method includes the steps of (a) obtaining sensitivity data for a collection of biological specimens, (b) obtaining gene expression data for the biological specimens in the collection, (c) constructing a model by a partial least squares method type 1 using said sensitivity data obtained in step (a) and at least a part of said gene expression data for the biological specimens obtained in step (b), and (d) making said model available to estimate the sensitivity of the biological specimen to the specific drug. The model estimates the sensitivity of a biological specimen to a specific drug.

The rejection of claim 1 contends that Rocke describes the construction of a model that estimates the sensitivity of the biological specimen to a specific drug by a partial least squares method type 1. Applicant respectfully disagrees.

In this regard, attention is respectfully directed to the description of the PLS family of methods in Martens. As can be seen, PLS-1 regression predicts a single y-variable from a block of X-variables, whereas PLS-2 repression predicts a whole block of Y-variables from a block of X-variables. *See, e.g., Martens*, page 308, para. 2; FIG. 4.

In contrast, Rocke describes that "Multi-class [Cancer] Classification" methods can be used to predict drug sensitivity of the tumor. *See, e.g., Rocke*, para. [0145]-[0146]. Rocke's "Multi-class Classification" is understood to be multivariate. *See, e.g., Rocke*, paras. [0053], [0056]. Rocke's multivariate "Multi-class Classification" is thus understood to use partial least squares method type 2. *See, e.g., Martens*, page 308, para. 2; FIG. 4 (describing the use of "several Y-variables" as "PLS2").

Accordingly, claim 1 is not anticipated by Rocke. Applicant respectfully requests that the rejections of claim 1 and its dependencies under 35 U.S.C. § 102 be withdrawn.

Rejections under 35 U.S.C. § 103

Claim 1 was rejected under 35 U.S.C. § 103(a) as obvious over Rocke and U.S. Patent Publication No. 2004/0199334 to Kovesdi et al. (hereinafter "Kovesdi").

The rejection contends that the subject matter recited in claim 1 would have been obvious to those of ordinary skill in light of Rocke and Kovesdi. Applicant respectfully disagrees. As discussed above, Rocke's multivariate "Multi-class Classification" is understood to use partial least squares method type 2 and would not have lead those of ordinary skill to construct a model

that estimates sensitivity of a biological specimen to a specific drug by a partial least squares method type 1 using sensitivity data for a collection of biological specimens and gene expression data for the biological specimens.

Kovesdi does nothing to remedy these deficiencies in Rocke. Indeed, Kovesdi relates to relating the structures and activities of chemical compounds. *See, e.g., Kovesdi*, Title. Kovesdi is thus not reasonably pertinent to either the present claims or to Rocke. There is no reason to believe that those of ordinary skill would find it obvious to turn to Kovesdi and constructing a model that estimates sensitivity of a biological specimen to a specific drug, as recited in claim 1.

Accordingly, claim 1 is not obvious over Rocke and Kovesdi. Applicant respectfully requests that the rejections of claim 1 and its dependencies under 35 U.S.C. § 103(a) be withdrawn.

It is believed that all of the pending claims have been addressed. However, the absence of a reply to a specific rejection, issue, or comment does not signify agreement with or concession of that rejection, issue, or comment. In addition, because the arguments made above may not be exhaustive, there may be reasons for patentability of any or all pending claims (or other claims) that have not been expressed. Finally, nothing in this paper should be construed as an intent to concede any issue with regard to any claim, except as specifically stated in this paper, and the amendment of any claim does not necessarily signify concession of unpatentability of the claim prior to its amendment.

Applicant hereby petitions that the period for response be extended by one month to include June 16, 2008. Please apply the excess claims fees, the fee for the one-month extension of time, and any other charges or credits to deposit account 06-1050.

Respectfully submitted,

Date: June 16, 2008

John F. Conroy
Reg. No. 45,485

Fish & Richardson P.C.
12390 El Camino Real
San Diego, California 92130
Telephone: (858) 678-5070
Facsimile: (877) 769-7945

JFC/jhg
10840897.doc

STATISTICAL PROCEDURES IN FOOD RESEARCH

Edited by

J. R. PIGGOTT
Food Science Division, Department of Bioscience and Biotechnology, University of Strathclyde, Glasgow, Scotland, UK

ELSEVIER APPLIED SCIENCE
LONDON and NEW YORK

ELSEVIER APPLIED SCIENCE PUBLISHERS LTD
Crown House, Linton Road, Barking, Essex IG11 8JU, England

Sole Distributor in the USA and Canada
ELSEVIER SCIENCE PUBLISHING CO., INC.
52 Vanderbilt Avenue, New York, NY 10017, USA

WITH 62 TABLES AND 113 ILLUSTRATIONS

British Library Cataloguing in Publication Data

Statistical procedures in food research.
1. Food industry and trade—Research—
Statistical methods
I. Piggott, J. R.
664'.0072 TP370.8

Library of Congress Cataloging-in-Publication Data

Statistical procedures in food research.

Bibliography: p.
Includes index.
1. Food—Research—Statistical methods. I. Piggott,
J. R. (John Raymond), 1950–
TX367.S73 1986 519.5'024664 86-16202

ISBN 1-85166-032-1

Phototypesetting by Interprint Limited, Malta
Printed in Great Britain by Galliard (Printers) Ltd, Great Yarmouth

PREFACE

The compilation of this book was in part prompted at a workshop on statistical methods at the 1984 Weurmann Symposium. After some confused and inconclusive discussion, it was clear that the participants in the workshop were not all familiar with the methods available. The participants wanted to learn more about them, but did not know where to start, since most of the available books were written for student or practising statisticians. A 'user's guide' to these methods was suggested as a solution, which would explain the background and provide examples of use and interpretation.

The methods described here owe much to the great statisticians, such as Pearson, Fisher, Hotelling, Bartlett and Wilks; some of the methods have seen major use in psychological and sociological research; cluster analysis has been developed as a major tool of numerical taxonomy; partial least squares regression has been considerably developed and used in chemometrics. Despite their disparate origins and development, these methods can all be used to help the food researcher to understand data—they are all tools for data analysis.

The wide availability and ease of use of computers and programs for multivariate analysis means that more researchers can use these methods, but this development is not an unmixed blessing; it means that the methods can be applied when they are not appropriate, by users who do not really understand them and uncritically accept the results.

I have asked the contributors to this book to try to explain the most common and most useful multivariate methods, after introducing the basic principles of statistical methodology and terminology. I hope the contents will be accessible to readers with little mathematical training, since the emphasis is on understanding the principles of the various methods, their applications and the interpretation of the results.

Chapter 9

PARTIAL LEAST SQUARES REGRESSION

M. MARTENS and H. MARTENS

Norwegian Food Research Institute, Ås, Norway

1. INTRODUCTION

1.1. Problems in Food Science and Technology

Food research and food production involve problems of biochemical, chemical, physical, technological, agronomical as well as sociological and psychological nature. Quality may be defined as the totality of features and characteristics of a product that bear on its ability to satisfy a given need (EOQC, 1976). In food research we want to understand the basic chemical and physical characteristics that cause our perception of quality. In the food industry we want to develop and maintain products of high quality from the producer to the consumer, also taking economical and political aspects into consideration, i.e. total quality control (Martens, M., 1984). For both problems achieving the final aim requires that we are able to relate 'products to persons'—relate hard fact instrumental measurements to more soft human data. In addition, high-speed instruments for quality control require conversion of non-selective data to selective information.

However, although there are thousands of variabilities in this two-block or multi-block space, biological systems are not infinitely complex: there exist systematic interrelationships reflecting the same basic phenomena (Fig. 1).

The *first* purpose of this chapter is to show that it is possible to extract the main information from various types of quality measurements by focusing the data onto a few, underlying latent variables.

1.2. Data Analytic Methods in Food Research

Two facts concerning food research measurements are clear:

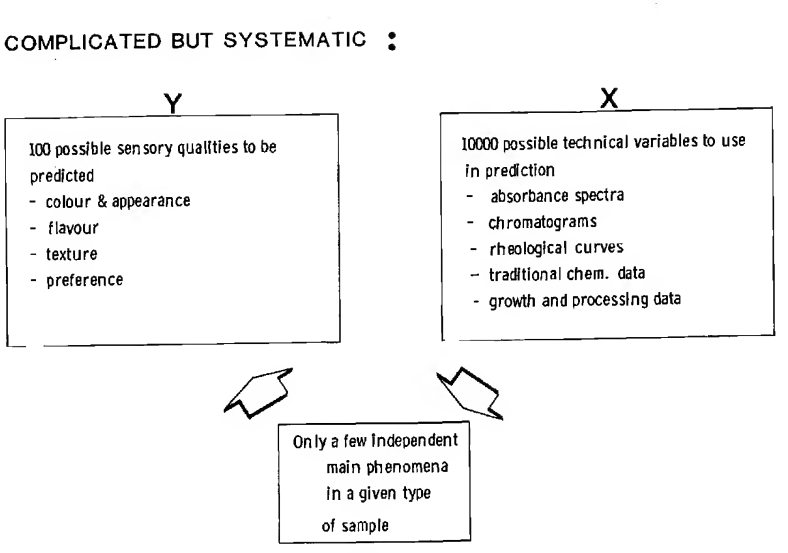


FIG. 1. A multitude of sensory as well as chemical, physical and optical measurements can reflect the same basic phenomena in a given type of biological material. Thus, e.g. 10 000 technical variables (**X**) might be used to predict 100 sensory variables (**Y**).

— We need *many* data to cover the complexity.
— We need data analytic tools so we, on one hand don't drown in data, and on the other hand, don't oversimplify by looking only at those data that we feel comfortable with, ignoring the rest, or lose information by averaging systematically opposing trends.

Data analysis involves fitting some sort of mathematical 'model' or 'formula' to the data by estimating certain 'model parameters', followed by an evaluation of the obtained parameters and the residual lack-of-fit:

$$\text{Data} = \text{'model' (with parameters)} + \text{residual} \qquad (1)$$

Any data analysis, be it simple or sophisticated, consists of:

(1) *A choice of data model*—a mathematical 'formula' expected to be suitable for the purpose of the experiment and the expected structure of the relationships and uncertainties in the data.
(2) A procedure for *fitting this model to the data*—for estimating the model's parameters. The procedure should correspond to the design of the experiment and should be based on reasonable assumptions about the uncertainties involved.
(3) *Validation* of the results obtained: Is the chosen model suitable for these data? Have we overfitted the data? Are there abnormalities (outliers) in the data? Are the results reasonable? The validation should not be performed on the data used for the fitting of the model to the data, unless special precautions are taken (e.g. cross-validation or leverage correction).
(4) *Presentation* of the data analytic results (parameters and residuals), for one's self during the data analysis and for others after the conclusions have been drawn.

Multivariate data analysis has increasingly been applied in food research during the last decade (Martens, M. and Harries, 1983). More popularly (Powers, 1984; Gacula and Singh, 1984; Martens, H. *et al.*, 1983c) as well as theoretically (Cooley and Lohnes, 1971) based descriptions of various methods have been given.

Multivariate data analytic techniques for finding the main information or structure in *one* data table (e.g. principal component analysis, factor analysis, cluster analysis) are described in other chapters of this book. However, the problems discussed in Section 1.1 also concern relationships between *two* or *more* data tables.

Thus, the *second* purpose of this chapter is to give a short overview of possible methods for quantitatively relating two or more blocks of data (Section 2), especially focusing on Partial Least Squares (PLS) regression (Sections 3, 4) and its properties on real data compared to other methods (Section 5).

1.3. What Does the Data Analysis Require from the Food Scientist?

Computerized multivariate data modelling can be of great help in food research and development. But there are many pitfalls in multivariate data analysis, at least in the way many standard statistical programs are used:

* An analytical chemist may be so fond of his favourite theory of what takes place in his samples that he forces a certain mathematical model onto his data and does not want to look at alternatives or error warnings. Examples: Linearized kinetic models of reactions, 'Beer's Law' models in spectrometry of mixtures.
He should first have used a 'soft' modelling technique like the PLS regression in order to let the data talk for themselves. He might then

have detected an obvious flaw in his theory! Or he may see that he only has one single observation of a certain experimental condition, and therefore must improve his experimental design before reliable conclusions can be drawn from the data.

* A statistician may receive excellent and informative data from the food scientist: A thousand different variables have been measured by a sophisticated orchestra of instruments like spectrophotometers and chromatographs. The food scientist wants to know how they correlate to sensory quality measured on the same samples. The statistician has to disregard all but a few of these thousand instrument variables, because his regression method cannot use more variables than objects. A lot of good data are wasted.
  Instead, the statistician or the food scientist himself should have used a method like the PLS regression, finding the main harmonies from the instrument orchestra, and using them for modelling the sensory quality. In that way the food scientist can learn about the different phenomena that affect his samples, seeing which variables give similar information, which give unique information and which give no information at all.

Traditionally, data analysis was more or less synonymous with statistical hypothesis testing. Except for the simplest types of *t*-tests and *F*-tests, hypothesis testing requires special expertise to ensure that the underlying assumptions of the test are fulfilled. But in food research today the main data analytic problem is not to *test* well defined hypotheses, but rather to *find* the information in a mass of data.

Gower (1983) discusses the division of responsibilities between the food scientist and the professional data analyst in more detail. All scientists should be familiar not only with the characteristics of their data but also with the models fitted and assumptions made.

Irrespective of data analytic technique, an evaluation of validity and practical relevance of the results is always required from the food scientist. Thus, the *third* purpose of this chapter is to discuss validation and interpretation procedures, especially concerning PLS regression (Section 6).

1.4. Short History of PLS

Partial Least Squares regression is a relatively new approach to multivariate data analysis. The basic concept was originally developed by the statistician Herman Wold and is described in Wold, H. (1982). The particular version of PLS regression to be focused on in this chapter is the Orthogonalized mode A two-block predictive PLS regression on latent variables (Wold, S. *et al.*, 1983a).

This 'PLS regression' has proven itself very useful both for calibration (Martens, H. and Jensen, 1983; Martens, H. *et al.*, 1983a,b; Wold, S. *et al.*, 1983a) and for general data interpretation (Wold, S. *et al.*, 1983b, 1984). It is a robust and intuitively appealing algorithm, rather than a theoretically derived method based on minimizing a certain statistical criterion. For this reason it has gained considerably more popularity among, for example, chemometricians than among statisticians, although even the latter group has given it some attention lately (Næs and Martens, H., 1985; Aastveit and Martens, H., 1986). Its mathematical relationship to matrix inversion is treated in Wold, S. *et al.* (1985).

Food research scientists should pay extra attention to the possibilities of relating different types of chemical, physical and biological as well as sensory and sociological information to each other. Thus, the *fourth* purpose of this chapter is to give a list of areas expressing PLS's potential for interdisciplinary applications (Section 7). Finally, a list of available PLS programs is given (Section 8).

2. OVERVIEW OF MULTIVARIATE TWO-BLOCK DATA ANALYTIC METHODS

2.1. Definitions and Distinctions

In the following, bold-face upper-case letters (e.g. **X**) refer to matrices (=tables=blocks of variables). Bold-face lower-case letters (e.g. **x**) refer to vectors (a column or row of data elements), and ordinary lower-case letters (e.g. x) refer to scalars (individual data elements=individual numbers). Upper-case ordinary letters are used to describe upper limits for running indices (e.g. $k=1,2,\ldots,K$).

The present chapter mainly concerns how to find quantitative relationships between two blocks of variables, **X** and **Y**.

Variable (=manifest variable) means input data from a certain analytic method, e.g. pH, absorbance at a certain wavelength, or sensory score for a certain flavour attribute. In this chapter a variable is represented by a *column* of numbers: by index $k=1,2,\ldots,K$ if it is represented in matrix **X**, and by index $j=1,2,\ldots,J$ if it is represented in matrix **Y**. Sometimes, K and J are replaced by p and q as the total number of variables in the **X**-block and **Y**-block, respectively.

Latent variable (=factor) means a 'variable' that is estimated during the data analysis in order to represent a general variability more or less common to a set of input variables directly or indirectly intercorrelated. It is represented by index $a=1,2,\ldots,A$.

Sample (=object) is used in its chemical sense of a single object, not in its statistical sense of a 'sampled subset of a population'. In this chapter a sample is represented by a *row* of numbers. It is represented by index $i=1.2,\ldots,I$. Sometimes, symbol I is replaced by n as the total number of samples.

Matrix means two-way table (samples × variables).

Vector means a row or column of data elements.

Model is mainly used to represent the formula or type of formula chosen to be fitted to the experimental data. The model consists of a certain combination of mathematical rules (e.g. conventional linear regression: $\mathbf{y}=b_0+\mathbf{x}b_1+\mathbf{e}$) between variables (here: **y** and **x**) and a set of model parameters to be estimated from the data (here: b_0, b_1). The estimation of the parameters also implies an estimation of the residual or lack-of-fit (here: **e**). Distributional assumptions about parameters and residuals are implicit parts of the model.

Parameters are the unknown 'numbers' in the model, to be estimated from fitting the model to the data. Note that parameters are not the same as constants; the term *constant* should in data analysis be reserved for true physical constants (e.g. Planck's constant).

Estimation is a procedure for finding the best possible values of the unknown parameters in a mathematical model. It is here used to mean the choice of method for computing the parameter values, plus the actual fitting of the mathematical model to a certain set of data. In this chapter the method of computing is always based on the principle of least squares residuals, but different least squares methods can be used for estimating the same parameter. Estimated values of parameters are represented by the 'hat' symbol, e.g. $\hat{\mathbf{b}}$, $\hat{\mathbf{y}}$, $\hat{\mathbf{e}}$.

Least Squares (LS) methods of fitting a mathematical model to a certain set of data involve minimizing the sum of the squared residuals, i.e. the squared differences between data and the corresponding predictions from the model. The fitting yields estimates for the parameters in the model. LS methods require a weighting of the relative reliability of the different residuals (different samples and/or variables) to be summed in multivariate LS methods.

Regression means fitting a mathematical model to a certain set of data by projecting some variables onto other variables, usually by Least Squares.

Calibration concerns how to find the mathematical formula that optimally converts, for example, instrument data **X** to results **Ŷ**. Once this formula has been determined (statistically 'estimated') the levels of the **Y**-variables can be predicted in new, unknown samples of the same kind, from their **X**-variables.

Prediction means to use the chosen model and the estimated model parameters to predict certain variables from other variables available. Example: after linear regression, predict **y** from the input data for variables **x** by: $\hat{\mathbf{y}}=\hat{b}_0+\mathbf{x}\hat{b}_1$.

Predictive methods are here used for regression methods that imply a prediction of certain variables **Y** from other available variables **X**, as opposed to '*correlative*' methods, in which no predictive direction is implied.

Classification also called 'supervised learning' or 'pattern recognition' (PARC), concerns the analysis of differences between classes of objects. Only classification followed by a quantitative prediction is relevant for this chapter (PARC level 3 and 4; Wold, S. *et al.*, 1984).

Validation is a procedure to ensure that the model and its estimated parameters really have predictive ability for future, unknown samples of the same general kind. Validation is an important step in data analysis and guards against overfitting, i.e. modelling just nonsense. The validation may be based on results from a new test set of samples not used in the actual model fitting. But the methods of *cross-validation* and *leverage correction* make it possible to use data from the same samples both for model fitting and validation (see Section 3.6).

2.1.1. Metric Data and Linear Additive Models

This chapter mainly concerns *metric* data analytic methods (assuming quantitative data at the 'ratio' (e.g. 4°K is twice as hot as 2°K) or 'interval' (e.g. 42 °F is 10 °F higher than 32 °F) measurement levels), as opposed to *non-metric* methods (accepting qualitative input data at 'ordinal' (cold < cool < hot) or 'nominal' (milk ≠ beer ≠ water) measurement levels); see Young (1981). But non-metric input data can be useful even as part of metric data sets, as some of the presented examples will show (Sections 4.3, 4.5, 5.2, 5.3 and 5.4).

Within psychometry different non-metric two-block multivariate data analytic techniques have been developed which may be of great interest for food scientists, e.g. a non-linear canonical correlation analysis (Burg and Leeuw, 1983) and non-metric regression analyses (Young, 1985). Non-metric multidimensional scaling methods are discussed in Chapter 6 in this book and in MacFie and Thomson (1984).

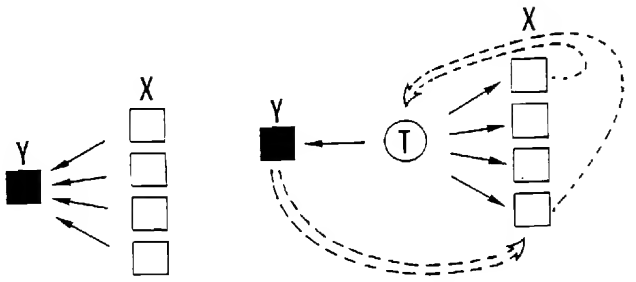


FIG. 2. Conceptual difference between (a) MLR and (b) PLS (here PLS1). In MLR, all the **X**-variables (here: 4) simultaneously model **Y** (here: only one **y**) (solid arrows). In PLS1, **Y** is used as a guide in the extraction of latent variables **T** from **X** (dotted arrows). These latent variables **T** (here: only one **t**) are in turn used for modelling both **X** and **Y** (solid arrows).

Further, this chapter mainly concerns *linear additive* mathematical models, as opposed to *non-linear* models. But by centring the data, linear additive models can approximate many types of non-linearities in data (Martens, H. *et al.*, 1983a; Geladi *et al.*, 1985).

2.1.2. Multicollinearity: Regression on Manifest or on Latent Variables? The multivariate two-block data analytic methods differ in one fundamental aspect of great practical importance (see Fig. 2): One family of methods, like Multiple Linear Regression (MLR) is based on regressions of **y** directly onto the individual **X**-variables (the 'manifest' variables). In the other family of methods based, for example, on the Partial Least Squares (PLS), the regression of **y** is done on so-called 'latent' variables, **T**, representing the main variation found to be common to many **X**-variables. Thus, multicollinearity among the **X**-variables gives a stability advantage instead of creating an instability problem. The two families are similar when the number of samples is very high. But when the number of samples is low, the more traditional methods can give graver estimation errors (see Section 5.1).

2.2. Regressing X on Y or Y on X, or both on Latent Variables?

One way of classifying the different methods for two-block modelling concerns the way irrelevant phenomena in the data are treated. By 'irrelevant phenomena' we here mean systematic types of variations in

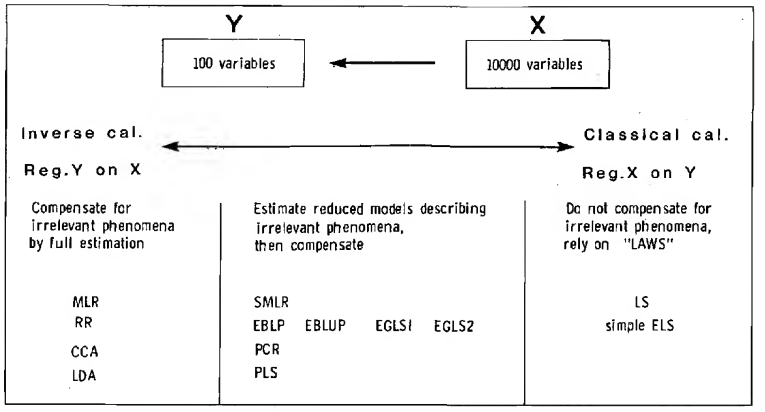


FIG. 3. Multivariate methods for relating two blocks of variables (e.g. **X** and **Y** from Fig. 1) to each other. Inverse versus classical methods distinguish between different treatments of irrelevant phenomena. Abbreviations: see text.

one data table that have no relationship to the variations in the other table.

Figure 3 positions a number of the two-block methods in this respect. In the univariate calibration literature a long debate has taken place on whether to regress **x** on **y** (classical calibration) or **y** on **x** (inverse calibration) in order to get the best predictions of **y** from future **x**-data. The conclusion from this univariate/bivariate field was that when the correlation between **x** and **y** was high, it did not matter very much. But if the correlation was low, then one should regress **y** on **x** if the calibration samples were representative for the distribution of future unknown samples, while one should regress **x** on **y** if they were not and one therefore might have to extrapolate outside the range of calibration. Noise in the **x**- and **y**-data complicates this conclusion.

This 'classical versus inverse' distinction has been extended to multivariate linear regression methods (Næs and Martens, H., 1984). Figure 3 uses this terminology to distinguish between different treatments of irrelevant phenomena.

2.2.1. Regressing **X** *on* **Y**-*variables*

Assume, for instance, that the available input variables $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots)$ from an instrument are affected by the con-

centration of the constituent to be determined, y_1, and by some well characterized interfering constituents, $y_2, y_3, \ldots$.

In *classical* modelling one relies on assumed physical 'laws' for how the **X**-data are generated from the **Y**-data. These methods have the ambition of explicitly modelling every *important phenomenon* $y_1, y_2, y_3, \ldots$ that affects the **X**-data in the samples:

$$\mathbf{X} = f(y_1, y_2, y_3, \ldots) + \mathbf{E} \quad (2)$$

where $f(\ )$ defines the mathematical model and **E** is the residual, ideally representing random measurement noise in **X**.

Thereby the well known interferences $y_2, y_3, \ldots$ can be described and compensated for, allowing determination of $\hat{y}_1$. If the correct physical model has been chosen, then this type of modelling gives the best results from a given set of data. But alas, in food science, such complete modelling is virtually never possible in anything but very pure model systems! The danger is therefore that irrelevant and/or unexpected phenomena $z_1, z_2, \ldots$ in **X** are ignored and never even checked for, with potential alias errors as result:

$$\mathbf{X} = f(y_1, y_2, y_3, \ldots, z_1, z_2, \ldots) + \mathbf{E} \quad (3)$$

The unmodelled phenomena $z_1, z_2, \ldots$ manifest themselves as apparent variations in the modelled phenomena $\hat{y}_1, \hat{y}_2, \hat{y}_3, \ldots$ leading to mistaken conclusions.

Least Squares (LS) curve fitting and the simplest types of Estimated Least Squares (ELS) regressions based on, for example, 'Beer's Law' and similar mixture models fall into this category of potential oversimplifications. Data models that cannot take into account unidentified covariance structures in the data are not recommended for anything but simple model experiments.

2.2.2. Regressing **Y** *on* **X***-variables*

On the other hand, in *inverse* modelling, the aim is to use the covariance structure within the data blocks to *compensate* for irrelevant variations in the samples analyzed, but without any attempts at modelling these irrelevant phenomena explicitly:

$$\mathbf{Y} = g(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots) + \mathbf{F} \quad (4)$$

where $g(\ )$ is the model and **F** the residual, ideally representing random measurement noise in **Y**. For example, MLR for one y-variable, this can be written

$$\mathbf{y} = b_0 + \mathbf{X}\mathbf{b} + \mathbf{f} \quad (5)$$

where b_0 and **b** are the regression coefficients (see Section 5.1.). If several **Y**-variables are to be modelled in terms of **X**, the MLR is called Multivariate Linear Regression.

The advantage of this approach over the classical one is that the user does not have to describe explicitly every phenomenon that affects the data; it is sufficient to ensure that all the important types of variations, including irrelevant phenomena, do vary in the data set used in the modelling. A disadvantage of this full inverse approach is the lack of explicit modelling of the various phenomena in the data; the resulting model shows how to avoid interferences, but not what the interferences are. A second, and in many practical situations graver disadvantage, is the problem of multicollinearity. With MLR, strongly intercorrelated **X**-variables create grave uncertainties in the modelling and interpretation, and corresponding decrease in prediction ability. Strong intercorrelations are often encountered when using instrument variables such as different wavelengths from a spectrophotometer, and will in addition always occur if the number of samples is low. In fact, MLR cannot be used at all if the number of samples is lower than the number of **X**-variables. Intercorrelations force the user to select a subset of **X**-variables and ignore the rest. This makes it impossible for the user to get a complete overview of the relationships in the data, and information is wasted.

This problem is caused by an unnatural and unnecessary mathematical assumption in these methods, namely that every regressor variable has unique information—it tells a story that is not being told by the others. Variables which only differ in their noise will be amplified grotesquely and force their noise nonsense into the final regression solution; this leads to numerical instability and statistical estimation uncertainties, as well as strong tendency for overfitting.

Attempts have been made at overcoming these estimation problems, for example, in MLR by artificially modifying the obtained covariance; this is called Ridge Regression (RR). If only the most dominant 'harmonies' (eigenvalues) in **X** are relevant for **Y**, then RR will work, but if intermediate and minor eigenvalues are important for **Y**, then RR will not be satisfactory, as has been shown for Near Infrared Reflectance (NIR) determination of protein in wheat (Fearn, 1983; Farebrother, 1984; Næs *et al.*, 1986).

Several other traditional multivariate methods like Canonical Correlation Analysis (CCA, see Section 5.2) and Linear Discriminant

Analysis (LDA, see Section 5.5) are based on the same assumption that each input variable has unique information. Hence these methods have the same limitations as MLR. They cannot be used directly unless the number of samples is much higher than the number of variables, and unless the variables are not otherwise too intercorrelated.

2.2.3. Intermediate Regression Methods

In between these two extremes we find a number of two-block regression methods designed to avoid the problems of the two extremes. Stepwise Multiple Linear Regression (SMLR) is a family of procedures to make MLR work by selecting only the 'best' **X**-variables. This may decrease the interpretability and represents a waste of information, unless one explicitly wants to find such a representative subset of **X**-variables to describe **Y**.

The remaining methods in this intermediate group have the ability that they both *describe* and *compensate for* irrelevant phenomena. Thus, for example, for predicting $\mathbf{y}_1$ from **X**, (e.g. via Eqn. (3)) both the identified interferents $\mathbf{y}_2$, $\mathbf{y}_3$, ... and the unidentified phenomena $\mathbf{z}_1, \mathbf{z}_2$, ... are modelled from the data.

The Estimated Best Linear Predictor (EBLP), the Estimated Best Linear Unbiased Predictor (EBLUP), and the two Estimated Generalized Least Squares predictors (EGLS1 and EGLS2) are similar. But they allow different distributional assumptions on whether or not the training set is representative for future samples, with respect to identified and unidentified phenomena in **X**. They rely on rather demanding matrix algebra, but have been shown to work well both from a theoretical (Brown, 1982; Næs, 1985a,b; Martens, H. and Næs, 1986; Sundberg, 1985) and practical (Skrede *et al.*, 1983) point of view.

An algebraically simple and versatile approach to multivariate regression methods is that based on 'soft modelling on latent variables' represented by Principal Component Regression (PCR) and PLS regression.

The latent variable methods can give a physical understanding like classical mixing models, while at the same time providing the statistical ability of the inverse regression methods of compensating for unidentified phenomena in the calibration data **X**.

This approach combines readings from all the relevant chemical or instrumental variables $\mathbf{X}=(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots)$ onto a few factors or latent variables that can be regarded as the main 'harmonies' between the **X**-variables. The intensities of these main harmonies in different samples are represented by a small factor score table $\mathbf{T}=(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_A)$ which models

both **X** and **Y**:

$$\mathbf{X}=f(\mathbf{T})+\mathbf{E}$$

and

$$\mathbf{Y}=g(\mathbf{T})+\mathbf{F} \tag{6}$$

where $f(\ )$ and $g(\ )$ define the factors mathematically. One may calibrate for a single y-variable or for a set of variables $\mathbf{Y}=(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots)$. Keeping to, for example, Eqn. (3), both the identified **X**-phenomena $(\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3, \ldots)$ and the unidentified **X**-phenomena $(\mathbf{z}_1, \mathbf{z}_2, \ldots)$ are modelled by the latent variables in **T**.

In practice an additive model is easiest to use. Here $f(\ )$ and $g(\ )$ are defined by estimates of so-called 'loading spectra' for each factor: Table $\mathbf{P}=(\mathbf{p}_1, \mathbf{p}_2, \ldots)$ defines the main harmonies in **X** and table $\mathbf{Q}=(\mathbf{q}_1, \mathbf{q}_2, \ldots)$ defines how their intensities **T** relate to the variables in **Y**:

$$\mathbf{X}=\mathbf{TP}+\mathbf{E}$$

and

$$\mathbf{Y}=\mathbf{TQ}+\mathbf{F} \tag{7}$$

Figure 4(a) shows this graphically for explicit modelling of only one y-variable. Figure 4(b) shows it for simultaneous calibration for J different **Y**-variables. Some sort of validation is used to select A, the number of factors optimal for the model.

Thus, soft multivariate calibration results in coefficients $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$, $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$. These can be used in the prediction of **Y** from **X** in future samples. This first maps the variation onto $\hat{\mathbf{T}}$ and $\hat{\mathbf{E}}$ and then predicts the **Y**-variables. Alternatively, this full spectral prediction can be simplified mathematically into a purely predictive model:

$$\hat{\mathbf{Y}}=\mathbf{X}\hat{\mathbf{B}} \tag{8}$$

where $\hat{\mathbf{B}}$ represents 'calibration coefficients' that directly predict **Y** from **X** without any estimation of scores **T** or residuals **E**. These $\hat{\mathbf{B}}$-coefficients are method-dependent functions of the calibration parameters $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$, $\hat{\mathbf{P}}$ and $\hat{\mathbf{Q}}$. In the limiting case they are identical to the MLR coefficients.

These two different modes for predicting **Y** from **X** give identical results; one yields better outlier detections, while the other is computationally faster.

The coefficients $\hat{\mathbf{B}}$ may also be regarded as a *rotation* of the factor axes $\hat{\mathbf{P}}$ to maximum prediction relevance: The A individual loading vectors $\hat{\mathbf{p}}_1$, $\hat{\mathbf{p}}_2$, ..., $\hat{\mathbf{p}}_A$ do not necessarily represent the individual 'spectra' or response-vectors of the various physical phenomena affecting the **X**-data; they may well represent combinations of the physical phenomena due to

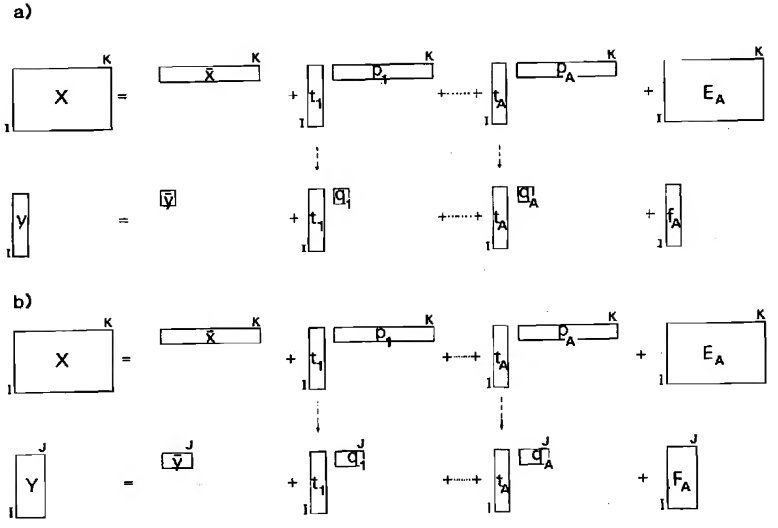


FIG. 4. (a) The data model of latent-variables regression methods for one individual y-variable (e.g. PCR, PLS1). (b) The corresponding data model for several Y-variables (PLS2).

intercorrelations between the phenomena in terms of amounts of 'spectra'. Columns $\hat{\mathbf{b}}_1, \hat{\mathbf{b}}_2, \ldots, \hat{\mathbf{b}}_j$ represent the directions in the **X**-space that correspond to the highest increase in $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_j$, respectively.

PCR and PLS here represent a larger family of latent-variable methods. There are several different methods of two-block regressions based on latent variables (Martens, H. and Næs, 1986): PCR, PLS regression, Fourier regression, Hruschka regression, latent root regression, etc. They differ mainly in the way the latent variables **T** are estimated as described above. PCR is the statistically most established method, and PLS regression may be considered as an extension of PCR.

3. PARTIAL LEAST SQUARES MODELLING: THEORY AND ALGORITHMS

3.1. What is Special about PLS?

Assessing the within and between block structures in, for example, two-block multivariate relationships can be quite complicated, because it

requires statistical description of a number of aspects concerning intercorrelations and mathematical rank, unique variances, etc. This is in statistics called the 'identification problem'. General programs like Jöreskog's LISREL (see Jöreskog and Wold, H., 1982) can be applied to such systems analysis, but they are difficult to use and sometimes have convergence problems. The PLS concept of H. Wold (1982) solves this identification problem in a simplified and user-friendly way.

The PLS regression is a systems analysis approach that optimizes several 'partial' (separate) sub-models, each by minimizing lack-of-fit residuals by the principle of least squares. Based on certain orthogonality properties, this allows a wide variety of modifications of the individual steps in the algorithm.

3.2. The PLS Family of Methods

Because of its flexibility, the PLS approach to systems analysis allows a number of different PLS algorithms to be developed to suit different analytical situations. Figure 5 shows some of the members of the PLS family already born.

The PLS family is here grouped first according to whether the methods are predictive or correlative.

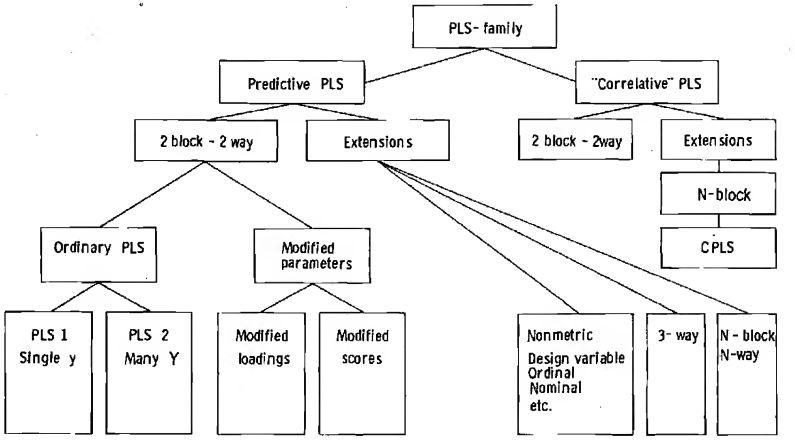


FIG. 5. The 'family' of Mode A PLS algorithm for 'soft modelling on latent variables.'

Among the predictive PLS methods we shall concentrate on analyses of two-block (**X**, **Y**) two-way data tables (samples × variables).

Among the two-block two-way methods we shall concentrate on the 'ordinary' PLS methods PLS1 and PLS2. PLS1 regression predicts a single **y**-variable from a block of **X**-variables, and thus resembles MLR. PLS2 regression predicts a whole block of **Y**-variables from a block of **X**-variables, and is thus akin to a predictive version of Canonical Correlation. These methods will be described in more detail below.

A range of Modified PLS versions have been published. Some of these concern modification of the factor scores with respect to additional information about the samples (Esbensen and Wold, S., 1983), others concern modifications of loading spectra by, for example, smoothing of continuous spectral data (see Martens, H. *et al.*, 1983b; Martens, H. and Næs, 1986). Preliminary tests have shown good results for Stepwise PLS, a loading modification designed to eliminate irrelevant **X**-variables with no information (Martens, H. and Næs, 1986).

In general the same types of modifications and extensions can be envisioned in both the predictive and the correlative groups. We shall in this chapter show the use of design variables and briefly mention the desirability of extensions like non-metric and three-way methods.

Concerning the correlative PLS methods we shall in the present paper only look at one, namely the new three-way or multiblock extension that we have termed Consensus PLS (CPLS).

3.3. The Predictive PLS1 Algorithm

PLS1 regression relates a single **y**-variable to a block of **X**-variables by focusing the **X**-variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots, \mathbf{x}_K$ onto a few factors $\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \ldots \hat{\mathbf{t}}_A$ ($A \ll K$) and using these estimated factors as regressors for the variable **y**. The general data model from Fig. 4(a) is described in more detail in this section.

The PLS1 algorithm is non-iterative and therefore very fast. It can be expressed in two apparently equivalent forms that give the same results for the orthogonal loadings and for prediction of **y**. The two algorithms (the *orthogonalized* and the *oblique*) were developed simultaneously in 1981 when it was discovered that the original two-block predictive PLS algorithm did not compensate for its non-orthogonal factor scores and therefore did not work on high-precision instrumental data.

3.3.1. The Orthogonalized PLS1 Algorithm

S. Wold (University of Umeå, Sweden) developed the orthogonalized algorithm that has been implemented in most PLS programs such as the SIMCA and the UNSCRAMBLER (see Section 8), and which is most easily modified. In this algorithm an extra set of loadings is estimated for the **X**-variables to ensure that the factor scores $\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \ldots, \hat{\mathbf{t}}_A$ are orthogonal to each other.

In order to understand the algorithm, consider a teaching process: In the data analysis, a set of variables $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots, \mathbf{x}_K$ are 'taught' how to reproduce **y** in a 'teaching' set of data. To ensure predictive ability, the training data should be accurate, representative and should not be overfitted!

This is parallel to a classroom situation:

The students $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \ldots, \mathbf{x}_K$ are taught how to reproduce the knowledge of the teacher **y** by training on certain didactic material that should be accurate and representative and that should not be misunderstood.

Assume that the class is to learn the main aspects of their country's history, in order to foresee the future development of their country. The teacher selects a history book, believed to be accurate and representative, and reads it aloud to the class. The teacher faithfully reads every sentence $i = 1, 2, 3, 4, \ldots, I$, including certain sentences with misprints.

The different students have different backgrounds and other previous knowledge on which they have to rely in order to understand what the teacher reads.

The analogue of MLR would be for the teacher to read the book once, and then rely on every student to tell something unique about the history. The first students will tell most of the important points. So, the last students are forced to try to make sense out of the misprint; their understanding of the main points is wasted.

In contrast, PLS1 regression attempts to give the best possible account of the history's main aspects, and tries to ignore the misprints: all the students are allowed to cooperate; what one student forgets others may remember, and what one student misunderstands others may correct.

The orthogonalized PLS1 algorithm goes like this:

ORTHOGONAL FACTORS ESTIMATED ONE AT A TIME (Fig. 4(a))

First, the individual **x**-variables are scaled by a weight to ensure that their variances correspond to their relative signal/noise levels. Then the

average for $\mathbf{y}$ and every variable $\mathbf{x}_k$ ($k=1, 2, \ldots, K$) in the I samples are calculated and subtracted, giving residual matrices: $\hat{\mathbf{E}}=\mathbf{X}-\mathbf{1}\bar{\mathbf{x}}$ and $\hat{\mathbf{f}}=\mathbf{y}-\mathbf{1}\bar{y}$. This centring of the data set is done in order to develop the subsequent multidimensional bilinear model around the estimated centre of the population of samples. This minimizes the effect of non-linearities, differences in offset, etc., and gives the best prediction ability if the training set of samples is representative for the future samples.

In the classroom picture each student is given a weight corresponding to how well he or she is expected to learn. The centring may correspond to ensuring that the teacher and the class concentrate on the history book and nothing else.

Then the parameters of the first factor $a=1$ are estimated (see below). The effect is subtracted to create new residuals $\hat{\mathbf{E}}=\hat{\mathbf{E}}_1$ and $\hat{\mathbf{f}}=\hat{\mathbf{f}}_1$, from which the second factor $a=2$ is estimated and subtracted. From residuals $\hat{\mathbf{E}}=\hat{\mathbf{E}}_2$ and $\hat{\mathbf{f}}=\hat{\mathbf{f}}_2$ the third factor $a=3$ is estimated, subtracted, etc. This process continues until no more valid factors can be obtained from the data.

In the classroom picture, the students pick up one aspect of the history each time the teacher reads. The teacher checks out what the students have understood and only repeats those parts of the history that have not yet been correctly reproduced. After a few such rounds, $a=1, 2, 3, \ldots, A$, all the significant points in the history may have been brought across.

The number of factors A cannot exceed $\min(K, I)$ for mathematical reasons; at that stage the X-matrix has been emptied for both information and noise, and consists of only zeroes. When $I>K$ so that MLR can be applied to the same data, then the limiting solution of $A=K$ is identical to the MLR solution when expressed in terms of $\hat{\mathbf{b}}$. As the example in Section 5.1 will show, in many practical cases this is carrying the modelling too far; the data become overfitted: minor noise phenomena and other nonsense have been drawn into the model and given apparent importance.

Overfitting the data is analogous to the over-zealous teacher going too far: After having ensured that the class has learnt all the important aspects of the history that they are mature enough to understand, he goes on trying to teach the class every little detail in the book. And the students will politely try to make some sense out of this nonsense.

Although not having understood, they will mumble something trivial that the teacher acknowledges as a sign that they still can learn more. In the end the teacher will even teach them the misprints in the book, and those students who never mastered spelling, will finally have their day.

A more careful teacher might have used 'history repeats itself', as validation criterion: new kings replace old kings. An incidental misprint should not be modelled; if for no other reason because it comes alone.

THE PARAMETERS OF EACH FACTOR ESTIMATED BY PARTIAL MODELS (Fig. 6(a))

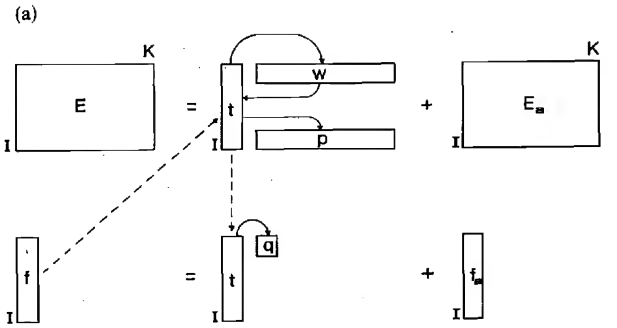
Using $\mathbf{y}$ as a guide in the estimation of the next loading-weight of $\mathbf{X}$:
The residual left in $\mathbf{X}$ and $\mathbf{y}$ after subtracting the effect of the previous factors are called $\hat{\mathbf{E}}$ and $\hat{\mathbf{f}}$. The remaining variation in $\hat{\mathbf{f}}$ is first used to ensure y-relevance of the new X-factor, $\mathbf{t}$: The $\hat{\mathbf{f}}$ is used as preliminary estimate for the new factor itself. Every variable in $\hat{\mathbf{E}}$ is projected onto this preliminary $\hat{\mathbf{t}}$, yielding preliminary estimates of the loading weight vector: $\hat{\mathbf{w}}=\hat{\mathbf{t}}'\,\hat{\mathbf{E}}/\hat{\mathbf{t}}'\,\hat{\mathbf{t}}$.

In the classroom picture, the teacher now repeats one more time those parts of the history that the students have not yet grasped. The students this time grasp a little more (as described by their loading weights). If the students are similar, they will all grasp the same points at the same time; if they are very different, they will grasp different things.

It is necessary to stabilize bilinear score × loading models against affine transformations of the type $2\times 6=3\times 4$. In PLS regression the conventional way of doing this has been to normalize this preliminary $\hat{\mathbf{w}}$ by a factor that ensures that the final $\hat{\mathbf{w}}$ has a square sum of one ($\hat{\mathbf{w}}\hat{\mathbf{w}}'=1\cdot 0$).

Estimating how different objects relate to this factor in $\mathbf{X}$:
The X-residual matrix $\hat{\mathbf{E}}$ is now projected onto the normalized loading weight vector $\hat{\mathbf{w}}$, to produce a score vector for the objects $i=1, 2, \ldots, I$: $\hat{\mathbf{t}}=\hat{\mathbf{E}}\hat{\mathbf{w}}'$.

In the classroom picture, all the students take a vote on what aspect of the story they should agree to account for now (as described by the factor score for each sentence). Ideally, the best students should carry stronger weight than the others. But even the best student can err, and



(b) Let $\hat{\mathbf{w}}_{21}$ and $\hat{\mathbf{W}}_{22}$ be the ordinary sample covariance estimators for cov(y, x) and cov(x). The first PLS vector $\hat{\mathbf{p}}_1$ is found by normalizing $\hat{\mathbf{w}}_{21}$ to unit length, i.e.

$$\hat{\mathbf{p}}_1 = \hat{\mathbf{w}}_{21}/\|\hat{\mathbf{w}}_{21}\|$$

and by induction

$$\hat{\mathbf{p}}_{a+1} \propto \sum_{m,n<a} \hat{q}_{am}\hat{\mathbf{p}}_n\hat{\mathbf{W}}_{22}\hat{\mathbf{p}}_m'\hat{\mathbf{p}}_m - \sum_{n<a} \hat{q}_{an}\hat{\mathbf{p}}_n\hat{\mathbf{W}}_{22} \quad \text{for } a \geqslant 1$$

$$\hat{\mathbf{q}}_a = (\hat{q}_{a1}, \ldots, \hat{q}_{aa})' = (\hat{\mathbf{P}}_a\hat{\mathbf{W}}_{22}\hat{\mathbf{P}}_a')^{-1}\hat{\mathbf{P}}_a\hat{\mathbf{w}}_{21}'$$

If $I<K$, we have matrix inversion problems (i.e. rank problems) when determining the $\hat{q}_a$ for $a>I$. We then define $\hat{q}_{am}=0$ and $\hat{p}_a=0$ if $a>I$.

(c) Principal Component Regression (PCR):

$$\hat{y} = \mathbf{x}\left[\mathbf{y}'\mathbf{X}\left(\hat{\mathbf{p}}_1\frac{1}{\lambda_1}\hat{\mathbf{p}}_1' + \cdots + \hat{\mathbf{p}}_A\frac{1}{\lambda_A}\hat{\mathbf{p}}_A'\right)\right]$$

PLS1 (Without Orthogonalized Scores):

$$\hat{y} = \mathbf{x}\left[\mathbf{y}'\mathbf{X}\left(\hat{\mathbf{I}}_1\frac{1}{\phi_1}\hat{\mathbf{I}}_1' + \cdots + \hat{\mathbf{I}}_A\frac{1}{\phi_A}\hat{\mathbf{I}}_A'\right)\right]$$

FIG. 6. The predictive PLS regression. (a) The conceptual PLS1 algorithm with orthogonal factor scores. (b) The PLS1 model. (c) PCR and PLS1 predictors have similar structure. (d) The conceptual PLS2 algorithm with orthogonal factor scores. (b) and (c) are discussed in Section 3.3.2. (d) is discussed in Section 3.4.

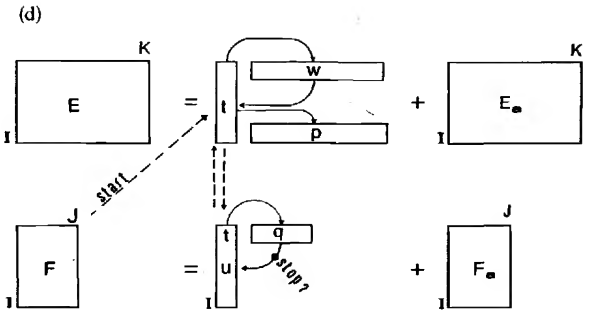


FIG. 6.—*contd.*

in the consensus he can still be over-ruled by a majority of other students.

Estimating how different X-variables relate to this factor:
In the orthogonalized PLS algorithm, $\hat{\mathbf{E}}$ is projected onto $\hat{\mathbf{t}}$, to ensure that subsequent factors' scores are orthogonal to this one. This gives a second loading vector: $\hat{\mathbf{p}}=\hat{\mathbf{t}}'\hat{\mathbf{E}}/\hat{\mathbf{t}}'\hat{\mathbf{t}}$. Then the effect of this factor is subtracted from **X**, creating new residuals $\hat{\mathbf{E}}$: $\hat{\mathbf{E}}_a=\hat{\mathbf{E}}_{a-1}-\hat{\mathbf{t}}_a\hat{\mathbf{p}}_a$.

In the classroom, all the students now accept the result of their vote, and make sure that everyone has understood this aspect of the story as well as they can (as described by the factor loading for each student). Disagreements among the students are for the time being postponed and kept as residuals in the back of their minds.

Estimating how the **y***-variable relates to this factor:*
Since the consecutive $\hat{\mathbf{t}}$-variables are orthogonal, it is sufficient to use univariate regression in fitting **y** (via $\hat{\mathbf{f}}$) to the scores from **X**. This gives the loading for **y**: $\hat{q}=\hat{\mathbf{t}}'\hat{\mathbf{f}}/\hat{\mathbf{t}}'\hat{\mathbf{t}}$. Then the effect of this factor $\hat{\mathbf{t}}$ is subtracted even from **y**, creating new residuals $\hat{\mathbf{f}}$: $\hat{\mathbf{f}}_a=\hat{\mathbf{f}}_{a-1}-\hat{\mathbf{t}}_a\hat{q}_a$.

Now the algorithm is ready for validation: deciding on whether to estimate one more factor or not (see Section 3.6).

In the classroom, the teacher listens to what the class has agreed about this time (as described by the factor loading for the teacher). He will later omit what was understood, correct what was misunderstood, and repeat what went unnoticed when reading it next time.

The teacher is now ready to decide on whether to read the remaining aspects of the history once more, or just accept the present performance as optimal for the given history and the given class.

The advantage of this orthogonalized PLS algorithm is its flexibility and its simplicity: Since $\hat{\mathbf{T}}'\hat{\mathbf{T}}$ is diagonal, it requires no complicated matrix inversions in the regression of y and $\mathbf{X}$ on $\hat{\mathbf{T}}$. A disadvantage for the user is the possible confusion of having two sets of X-loadings, $\hat{\mathbf{W}}$ and $\hat{\mathbf{P}}$, that may or may not be very similar. $\hat{\mathbf{W}}$ is orthonormal and shows the covariance structure in $\mathbf{X}$ relevant to y, while $\hat{\mathbf{P}}$ is not necessarily orthogonal and shows the intra-$\mathbf{X}$ covariance structure of the $\mathbf{X}$-variables relative to $\hat{\mathbf{T}}$. If the main intra-$\mathbf{X}$ covariance phenomena are also the ones relevant for y, then $\hat{\mathbf{W}}$ and $\hat{\mathbf{P}}$ are similar to each other (and to the eigenvectors of $\mathbf{X}'\mathbf{X}$; in such cases PLS1 and PCR give the same results).

3.3.2. The Oblique PLS1 Algorithm

H. Martens (NINF, Norway) developed the other algorithm for PLS1, which simply corrects for the problem of intercorrelated factor scores instead of eliminating it. The method was used in the calibration of fluorescence spectra for botanical components in wheat flour (Jensen *et al.*, 1982; Jensen and Martens, H., 1983).

It is slightly more complicated to program since it involves a matrix inversion of non-diagonal $\hat{\mathbf{T}}'\hat{\mathbf{T}}$ for the multiple linear regression of y on the non-orthogonal factor scores $\hat{\mathbf{T}}$. But on the other hand, it has fewer steps and is easier to study from a theoretical point of view. It yields only one set of loading vectors, $\hat{\mathbf{P}}$, which is identical to $\hat{\mathbf{W}}$ in the orthogonalized PLS1.

Næs and Martens, H. (1985) give a description of the PLS1 regression from a statistical point of view. Their study is based on the oblique PLS algorithm, and their description of the PLS1 model is given in Fig. 6(b). In Fig. 6(c) they show that the PLS1 predictor can be expressed in a way quite similar to the more well known PCR: ignoring for simplicity the averages, the formula that predicts y from the $\mathbf{x}$-vector of a new sample is a product of this new $\mathbf{x}$-vector and a function of the cross product of the calibration data y and $\mathbf{X}$, times a sum of a weighted contribution from different eigenvectors. In PCR these are the eigenvectors of $\mathbf{X}'\mathbf{X}$ itself, $\mathbf{p}_a$, $a=1,\ldots,A$. The weights are the inverse of the corresponding eigenvalues λ_a. (PCR with $A=K$ corresponds to MLR, which shows why MLR gives problems when $\mathbf{X}$ is multicollinear; this implies some eigenvalues λ_a close to zero, and division by near zeroes inflates $\hat{\mathbf{b}}$ and is detrimental to the prediction ability.)

In PLS1 it is the eigenvectors and eigenvalues of a *function* of $\mathbf{X}$ that enter into the predictor, but otherwise the two methods are similar. In Fig. 6(c) the vectors $\hat{\mathbf{I}}_1,\ldots,\hat{\mathbf{I}}_a$ are the eigenvectors of $\mathbf{X}'\mathbf{X}$ projected onto the PLS space $\mathbf{P}$, and $\phi_1, \ldots, \phi_A$ are the corresponding eigenvalues. In practice the authors could verify theoretical considerations that showed the PCR and PLS1 to be relatively similar. But PLS1 in general gave simpler models (fewer factors, A) and equal or better predictions than PCR, apparently since the use of $\hat{\mathbf{f}}$ in the estimation of $\hat{\mathbf{p}}$ ensures that $\mathbf{X}$-variations relevant to y are estimated before $\mathbf{X}$-variations irrelevant to y.

3.4. The Predictive PLS2 Algorithm

The PLS2 algorithm gives a predictive modelling of a whole block consisting of J different $\mathbf{Y}$-variables from a block of $\mathbf{X}$-variables, as illustrated generally in Fig. 4(b). It is most easily implemented as an extension of the orthogonalized PLS1 algorithm. Figure 6(d) shows that the algorithm, in contrast to PLS1, is iterative for each factor. Convergence is apparently no problem in the algorithm.

For a given factor number a, the preliminary loadings $\mathbf{w}$ are estimated as in orthogonalized PLS1 by regressing $\hat{\mathbf{E}}$ on some column of starting values from $\hat{\mathbf{F}}$.

As before, score vector $\mathbf{t}$ for a certain factor is estimated for the $\mathbf{X}$-variables by regressing $\hat{\mathbf{E}}$ on normalized $\hat{\mathbf{w}}$. The $\mathbf{Y}$-variables are then modelled by regressing their residuals $\hat{\mathbf{F}}$ on the estimated scores. Since we have more than one $\mathbf{Y}$-variable and hence get a whole $\hat{\mathbf{q}}$-vector instead of the single $\hat{q}$ element for each factor in PLS1, it is possible to obtain a preliminary estimate of the factor scores $\hat{\mathbf{u}}$ from the $\mathbf{Y}$-variables: by regressing the old $\mathbf{Y}$-residuals $\hat{\mathbf{F}}$ on the $\mathbf{Y}$-loadings $\hat{\mathbf{q}}$ we get $\mathbf{Y}$-scores $\hat{\mathbf{u}}=\hat{\mathbf{F}}\hat{\mathbf{q}}'/\hat{\mathbf{q}}\hat{\mathbf{q}}'$.

In order for the final $\mathbf{X}$-scores $\hat{\mathbf{t}}$ to model the largest remaining phenomenon in the $\mathbf{Y}$-variables, the PLS2 algorithm for each factor iterates back and forth between modelling $\hat{\mathbf{E}}$ from $\hat{\mathbf{u}}$ (the main variation in $\hat{\mathbf{F}}$) by $\hat{\mathbf{w}}=\hat{\mathbf{u}}'\hat{\mathbf{E}}/\hat{\mathbf{u}}'\hat{\mathbf{u}}$ and modelling $\hat{\mathbf{F}}$ from $\hat{\mathbf{t}}'$ (the main variation in $\hat{\mathbf{E}}$) by $\hat{\mathbf{q}}=\hat{\mathbf{t}}'\hat{\mathbf{F}}/\hat{\mathbf{t}}'\hat{\mathbf{t}}$. In the final solution, $\hat{\mathbf{t}}$ defines the factor while $\hat{\mathbf{u}}$ only represents a temporary estimation tool in this predictive version of PLS regression.

In the classroom picture we have J different teachers $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_J$

teaching the history in slightly different ways to the students $\mathbf{x}_1$, $\mathbf{x}_2, \ldots, \mathbf{x}_K$. The students first try to grasp those aspects that many of the teachers seem to agree upon. This probably gives the most important aspects of the history. Once that has been understood, the students go for the peculiarities of the individual teachers. The class will eventually end up trying to mimic their mistakes if no one tells the class what is significant and what is not.

3.5. Consensus PLS

The consensus PLS (CPLS) aims at relating several data-matrices to each other. For instance, in sensory descriptive analysis: What is the 'result' from a panel of several sensory judges? If half of the judges misunderstand the instructions and systematically give score 9 when the others give score 1 and vice versa, a simple averaging over all the judges may result in zero information. Confusion of terms likewise reduces information if simple averaging is used.

Another problem also has to be dealt with if several sensory attributes have been evaluated: Even the average or consensus of the sensory panel is an $I \times K$ table that often is too large to be interpreted directly. Further multivariate data compression is needed in order to focus the attention on the main reliable information in the consensus table, summarizing redundancy and leaving out random noise.

With CPLS we want to describe the samples and the judges as parsimoniously as possible by interrelating the main systematic differences between the judges, instead of just averaging over them. The CPLS algorithm allows that by combining a translation/rotation/scaling of Procrustes Rotation type (see Chapter 7 in this book) with a PCA-like rank reducing data compression in one single algorithm.

The correlative three-way CPLS algorithm was recently developed by Wold, S. *et al.* (1986). For simplicity it will only be illustrated graphically here. Figure 7(a) shows the basic data model exemplified for analysis of sensory panel data: I samples or objects have been assessed by J judges, each judge evaluating each sample with respect to K sensory terms. Thus the input data is the $I \times J \times K$ three-way matrix $\mathbf{X} = (x_{ijk})$.

The mean of every judge's evaluation on every term, $\bar{\mathbf{X}} = (\bar{x}_{jk})$, is first subtracted. A succession of orthogonal consensus factors $\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \ldots, \hat{\mathbf{t}}_A$ are then extracted in order to account for as much as possible of the systematic variance in $\mathbf{X}$. These score vectors are the columns of $\hat{\mathbf{T}}$ in the figure.

Each factor score vector $\hat{\mathbf{t}} = (\hat{t}_{ia})$ is a weighted aggregate of the

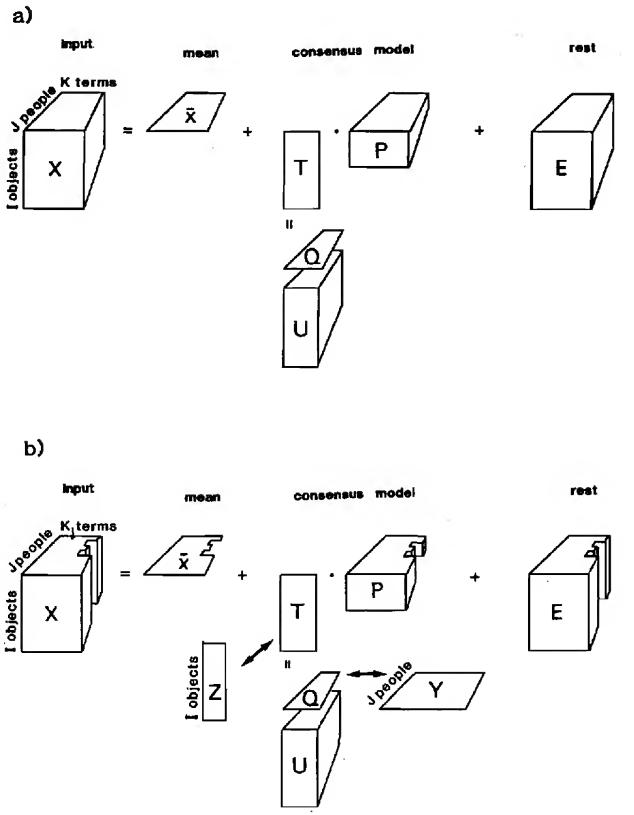


FIG. 7. The conceptual CPLS data model. (a) The parameters $\bar{\mathbf{X}}$, $\mathbf{T}$, $\mathbf{P}$ and $\mathbf{E}$ are extensions of those in Fig. 4. (b) CPLS has been modified compared to (a) to illustrate that the CPLS can handle input data when the J different people have not evaluated the same terms or even the same number of terms, K_j. Further extensions of the CPLS algorithm to incorporate additional information for interpretation, e.g. certain independent information $\mathbf{Y}$ about the J people, and certain independent information $\mathbf{Z}$ about the I objects.

individual judges' score of the object with respect to this factor number a: $\hat{\mathbf{U}}=(\hat{u}_{ija})$. The weighting vector $\hat{\mathbf{q}}=(\hat{q}_{ja})$ defines this weighted aggregate: $\hat{\mathbf{t}}=\hat{\mathbf{U}}\hat{\mathbf{q}}$.

Each factor has a loading matrix $\hat{\mathbf{P}}=(\hat{p}_{jka})$ which shows how each person, j, and each term, k, relates to this factor number a. The product of $\hat{\mathbf{T}}$ and $\hat{\mathbf{P}}$ defines the estimated data model for $\mathbf{X}$, from which the residual matrix $\hat{\mathbf{E}}$ is defined. Using the analogy with two-way matrix algebra, we can write:

$$\hat{\mathbf{E}}=\mathbf{X}-\bar{\mathbf{X}}-\hat{\mathbf{T}}\hat{\mathbf{P}} \tag{9}$$

Ordinary PLS validation criteria can be used for determining the number of significant factors. In the dataset to be presented later in Section 4.6, we used a leverage correction of the estimated residual variance of each variable x_{jk} to guard against overfitting. Figure 7(b) shows how the CPLS results can be related to external information about the I objects (**Z**-block) or the J people (**Y**-block).

3.6. Validation Procedures

The validation can be done in different ways for all the algorithms described above, in the following discussed with respect to PLS1.

Testing on independent validation sets requires extra, independent samples with known **X**- and y-data for which s^2, the mean squared error of $\hat{\mathbf{f}}$, is computed. The underlying requirement is that variation types or phenomena being modelled in the training set should also be relevant for the test set. This validation approach is somewhat wasteful, since the test data are not used in the actual modelling. The next two methods make it possible to use data from the same samples both for model fitting and validation.

Cross-validation is a procedure that repeats the modelling several times, each time using only part of the training samples in the model estimation and the rest as test samples from which y is predicted from **X**. After having used all the samples in turn as test samples, an estimate of the prediction error, s^2, is obtained as their mean squared error of $\hat{\mathbf{f}}$. Thus, its underlying requirement is that every phenomenon being modelled must be present in several samples. This method works well but is computationally somewhat time consuming.

Leverage correction is a new approach for directly converting the residuals $\hat{\mathbf{f}}=\mathbf{y}-\hat{\mathbf{y}}$ of the training set used for model estimation to estimates of the corresponding prediction error. The method is described for MLR in, for example, Cook and Weisberg (1982) and adapted for PLS regression by Martens, H. and Næs (1986).

Its underlying requirement is that even phenomena being modelled should be detectable in at least two samples in the training set in order to be accepted as valid. The leverage correction of a sample's residual $\hat{f}$ is $\hat{f}_i/(1-h_i)$ where the leverage h_i is approximated by:

$$h_i=\sum_{a=1}^{a} t_{i\alpha}^2/(\mathbf{t}_\alpha'\,\mathbf{t}_\alpha) \tag{10}$$

The leverage-corrected prediction variance for $\hat{\mathbf{y}}$ used here is:

$$s^2=\sum_{i=1}^{I}(\hat{f}_i/(1-h_i))^2/(I-A-1) \tag{11}$$

(The degrees-of-freedom correction $(I-A-1)$ is included in order to compensate for the influence exerted by $\hat{y}_{ij}$ on the estimation of **W**.)

To our knowledge the present paper is the first time leverage correction has been used in practice for PLS regression; and in our experience it gives results identical to or very similar to the cross-validation, although with considerably less computation time.

Irrespective validation method, new factors $a=1,2,\ldots,A$ may be accepted as long as s^2 continues to decrease (unless common sense indicates that systematic data errors are being modelled) as described in more detail in Section 6.2.

At the dimensionality A, corresponding to the minimal prediction error s^2, the PLS model consisting of $\hat{\mathbf{W}}$, $\hat{\mathbf{P}}$ and $\hat{\mathbf{q}}$ can be expressed in terms of the y-rotation, $\hat{\mathbf{b}}$:

$$\hat{\mathbf{b}}=\hat{\mathbf{W}}'\,(\hat{\mathbf{P}}\hat{\mathbf{W}}')^{-1}\hat{\mathbf{q}} \tag{12}$$

These estimates $(\hat{\mathbf{W}}, \hat{\mathbf{P}}, \hat{\mathbf{q}}, \hat{\mathbf{b}})$ obtained from the validated A-dimensional solution procedure should be used to present the optimal PLS model (see Section 6).

4. APPLICATIONS OF PLS REGRESSION

4.1. Overview of the Examples

4.1.1. Concerning the Measurements

Some of the principles discussed above are to be illustrated by recent work at the Norwegian Food Research Institute (NINF). The examples aim at relating two or more blocks of variables with at least one of the blocks containing data from sensory analysis; see Piggott (1984) for basic

literature in sensory analysis. The input data represent various types of quality measurements.

Descriptive sensory analysis using a panel of 12 trained judges under controlled conditions at the NINF's sensory laboratory. Each sensory attribute was evaluated along a 1–9 point intensity scale (1 = low; 9 = high intensity); the sensory data being input directly through a microcomputer system (Martens, M., 1985).

Consumer testing using about 100 untrained consumers at a central location giving preference scores along a 1–9 point hedonic scale (1 = dislike extremely; 9 = like extremely).

Instrumental methods which here means either traditional chemical and physical measurements or data from modern rapid spectrophotometers.

Product registrations including information about the origin of the product, or processing parameters, e.g. agronomical yield, growth location or temperature data.

4.1.2. Concerning the Data Analysis

Some details concerning the data analytic treatment were common for all the examples. The following list may serve as a model for information necessary to give from PLS-analysis.

Input data consisted of mean values across replicates where nothing else is mentioned.

Missing values: In some data sets a few reasonably well distributed missing values in the input data were handled by the software by ignoring these data elements in the PLS algorithm steps where they occurred.

Weighting of variables: In most of the predictive PLS analyses each variable was standardized to unit variance prior to the multivariate analyses in order to ensure each variable an equal chance of influencing the modelling. In the CPLS analysis the input data were analyzed directly.

Weighting of objects: All the objects were given the same weight since we assume the same analytical precision for each of them.

Design variables, often called dummy variables, here refer to variables with values 1 or 0, identifying different objects with respect to, for example, season, site and process parameter.

Validation criteria: Cross-validation and leverage-corrected residuals were used as well as interpretability of the solutions (defined in Sections 3.6 and 6).

4.1.3. Presentation of PLS Model Results

Factor loadings: $\hat{\mathbf{p}}_k$ (for the **X**-variables $k = 1, 2, \ldots, K$) and $\hat{\mathbf{q}}_j$ (for the **Y**-variables $j = 1, 2, \ldots, J$) give the direction of each factor through the object space spanned by the **X**- and **Y**-variables. A non-zero loading means that the variable's variation is correlated to and hence more or less described by this factor. Positive and negative loadings refer to positive or negative relationships to the factor (see Fig. 8).

Factor scores: $\hat{\mathbf{t}}_i$ gives the position of each object, i, on the obtained factor axis, i.e. the distance from the average or centre of the training set of the projection point of the object on the factor axis. A non-zero factor score means that the object's deviation from the average to some degree is described by this factor. Positive and negative scores refer to positive or negative relationships to the factor (see Fig. 8).

Rest variances express conceptually what variation is left unmodelled for the different **Y** (or **X**) variables by the PLS-analysis after a certain number of factors. More statistically it is here the mean square of the

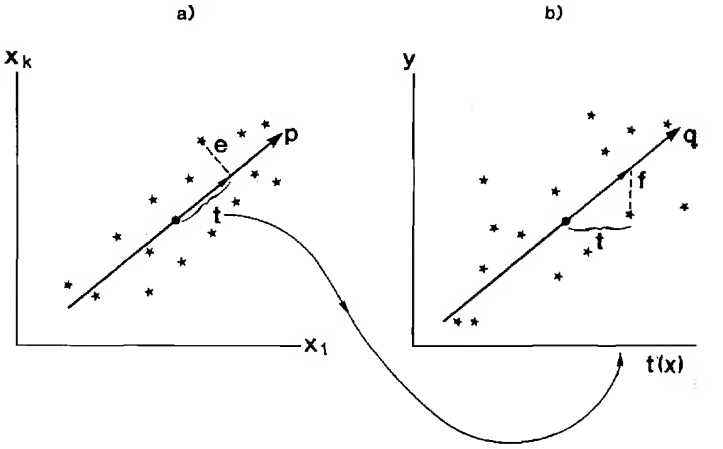


FIG. 8. The geometrical meaning of the PLS parameters. (a) Objects are represented by points in the K-dimensional **X**-space shown for two of these **X**-variables. The loading vector $\hat{\mathbf{p}}$ of a factor gives its direction in this **X**-space, its scores $\hat{\mathbf{t}}$ give the position of the objects, projected down on the factor relative to the origin (the average of objects). The residual vector, $\hat{\mathbf{e}}_i$, gives the lack-of-fit of each object to the factor axis. (b) Each score vector, $\hat{\mathbf{t}}$, is in turn used as regressor for the **Y**-variables, giving $\hat{\mathbf{q}}$ and $\hat{\mathbf{f}}$.

deviation **f** (or **e**) (see Fig. 8) of each variable. Its square root is also called standard error of prediction or prediction error (see Section 3.6). In most of the examples leverage-corrected residuals are given as the rest variance as a percentage of the total prediction variance.

Percent explained variance is (100-percent rest variance).

Rotation of the PLS-solution other than $\hat{\mathbf{b}}$ is not done due to the clear interpretability of plots of $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$.

4.2. PLS1 with One y-variable and Many X-variables

Example: Predict consumer preference of peas (*Pisum sativum* L.) from descriptive sensory analysis.

Model: In the PLS1 in Fig. 4(a), model $J=1$ preference variable **y** from $K=11$ sensory variables (**X**-block) over $I=16$ objects.

Experimental: The preference variable was obtained by averaging the

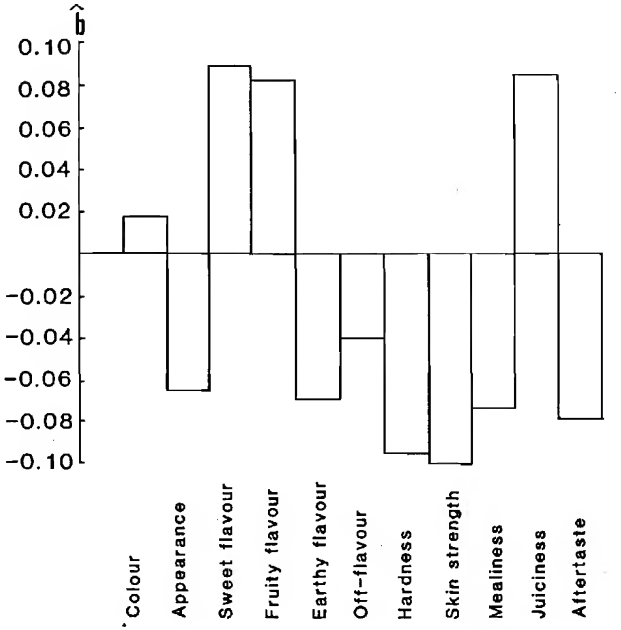


FIG. 9. Estimated regression coefficients $\hat{\mathbf{b}}$ for predicting consumer preference from $\mathbf{X}=11$ sensory descriptive attributes calculated for PLS1 using one factor.

data from 96 consumers for each of the 16 pea samples. This was done here to give an example of predictive modelling by PLS1. But for the purpose of understanding consumers' response, the preference data were also treated individually (see Section 4.5) as well as grouped according to demographic variables (reported elsewhere). The 11 sensory attributes described in Fig. 9 were evaluated in three replicates. Details about sensory profiling of the same pea batches are given in Martens, M. (1986b).

The 16 pea samples were blanched, frozen and analyzed within one month from harvesting in 1984. They were selected to represent relevant variation with respect to maturity levels and varieties.

Results and discussion: About 73% of the total leverage-corrected variance in preference could be explained by the sensory descriptive variables (**X**-block) after three PLS1 factors. The first and main factor related high preference to the sweet, fruity, juicy, not earthy and no aftertaste attributes. To focus on which variable was best at predicting **y**, the estimated regression coefficients, $\hat{\mathbf{b}}$, obtained from the PLS1 modelling were calculated as seen in Fig. 9. Strongly non-zero $\hat{b}_k$ for an **x**-variable means high predictive relevance. From this it seems clear that both flavour and texture variables and to a less extent colour, are important for pea preference.

The practical consequences of these results may be illustrated by an example (Fig. 10): some results from the descriptive sensory analysis for three of the pea samples and evaluation of the same samples by the consumer plotted together, made it possible to see how these tests were related. The sample profiled to be sweet and fruity and less hard than the others, was liked the most (hedonic score=6·7). The instrumentally measured tenderometer value (TV), not used in the PLS1 modelling, but conventionally used to express maturity of peas, was obtained for interpretation, and showed that this preferred pea (TV=102) was relatively immature.

Conclusions: The PLS1 regression indicated that sensory descriptive analysis of frozen green peas to some degree can predict an average consumer response; about 73% of the variation in the consumers' preference could be explained by a sensory profile panel by the bilinear modelling. The many PLS parameters $\hat{\mathbf{W}}$, $\hat{\mathbf{P}}$ and $\hat{\mathbf{q}}$ could be summarized by a single $\hat{\mathbf{b}}$-vector.

4.3. PLS2 with Many Y-variables and Many X-variables

Example: Explore relationships between analytic sensory and chemical

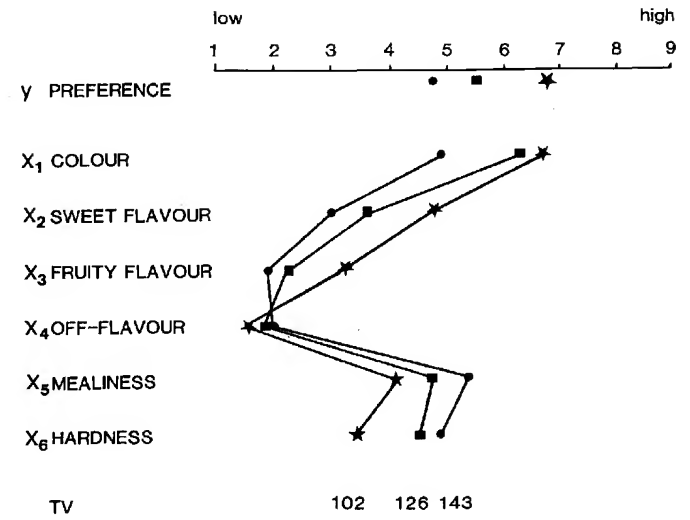


FIG. 10. Example of laboratory sensory analysis used to predict consumer preference. Three pea samples are profiled (1–9 point scale) with respect to some of the X-attributes from Fig. 9, and related to consumer preference (y). The 'star'-line represents a sensory description of a well liked pea. A maturity index (TV) is given for the same samples.

measurements of swedes (*Brassica napus* var. *napobrassica* L.) from different seasons and growth-sites.

Model: In the PLS2 in Fig. 4(b), model $J=10$ sensory variables (**Y**-block) from seven chemical variables + six design variables, in total $K=13$ (**X**-block) over $I=46$ objects.

Experimental: The 10 sensory attributes (total colour, crispness, chewing resistance, juiciness, total texture, fruity-, sweet-, bitter-, and sulphurous flavour, total flavour) were evaluated for each sample in three replicates. The seven chemical measurements (soluble solids, dry matter, total titratable acids, pH, sucrose, glucose and fructose) were performed in two replicates and calculated as a percentage of fresh weight. Details about the sensory and chemical analyses are given elsewhere (Fjeldsenden *et al.*, 1981; Martens, M. *et al.*, 1983a).

The 46 raw swede samples were chosen to span the relevant range of variation with respect to swedes bound for the Norwegian market. Three different growth seasons (1978, 1979 and 1980) and three different sites

(east-, west-, north-Norway) constituted the six design variables; for simplicity no interaction variables were included in the example.

Results and discussion: Leverage-corrected residuals showed optimal prediction ability for four PLS regression factors, explaining about 46% of the total variance in **Y** (Fig. 11). The prediction error variance is plotted as a function of increasing numbers of PLS factors. The prediction error reached a certain minimum (i.e. lowest rest variance) in **Y**, before increasing again with increasing model complexity. This general phenomenon is discussed in Section 6 (Fig. 27(a)). The corresponding **X**-residuals showed a continual decrease, ending at zero residual after 13 PLS factors, i.e. the MLR solution. This MLR solution has no predictive ability for **Y**. Figure 11 gives a good illustration of the PLS philosophy that a maximum of variation in **X** is drawn out to describe **Y** with a minimum risk of modelling just noise in **X** and **Y**.

PLS factors 1 versus 2 and 1 versus 3 are respectively shown for the loadings in Fig. 12(a,b) and for the scores in Fig. 13(a,b). Factor 1 constituted a relationship between, on one hand, the flavour variables, especially sweet (−) but also bitter (+), and, on the other hand, the

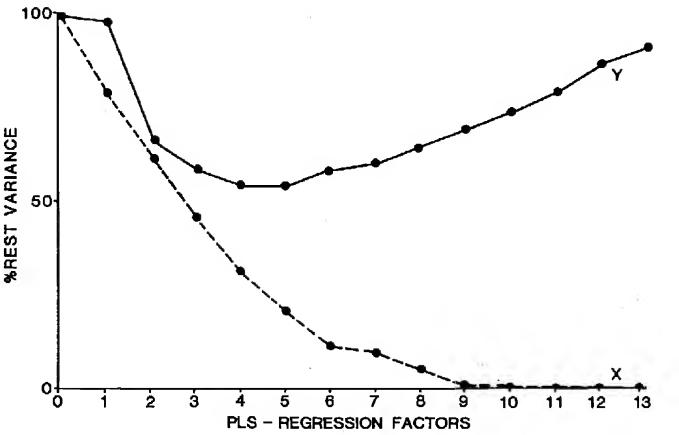


FIG. 11. Percentage rest variance after each of 13 factors for the **Y**-block (10 sensory variables) and the **X**-block (13 chemical and design variables) from a PLS2 analysis of the 46 swede samples. Leverage-corrected variances for both **X** and **Y** are given. The curved behaviour of the prediction error in **Y**, is discussed in Fig. 27.

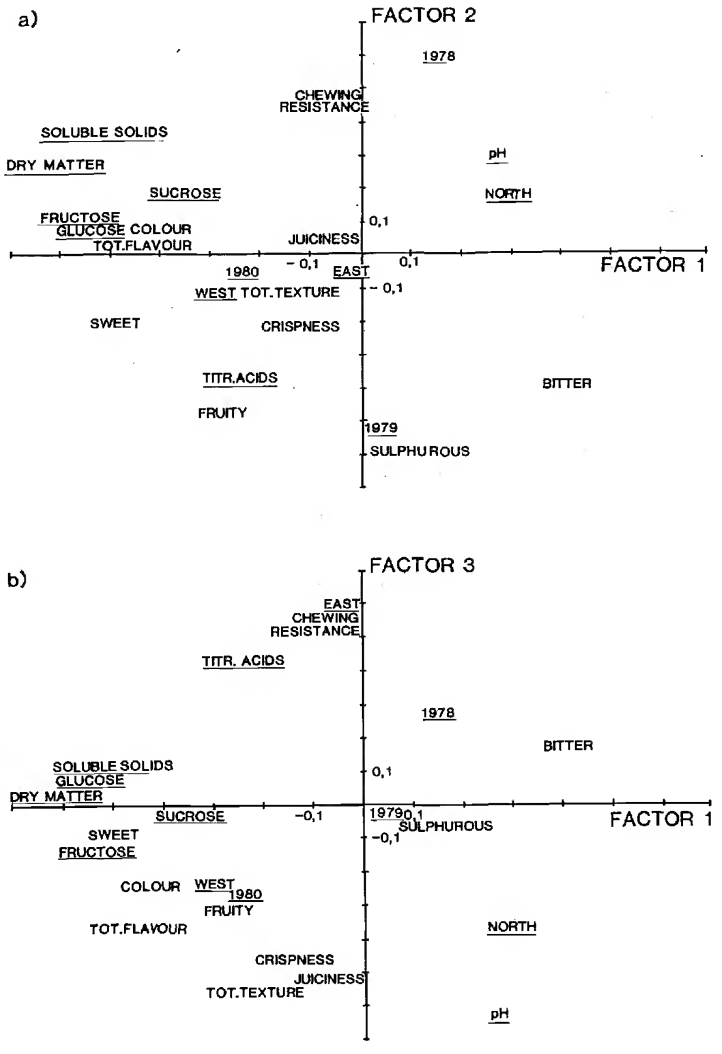


FIG. 12. PLS2 loadings for the 10 sensory variables (Y-block), seven chemical and six design variables (X-block) analysed on the 46 swede samples treated as in Fig. 11. (a) Factors 1 and 2. (b) Factors 1 and 3. The X-variables are underlined.

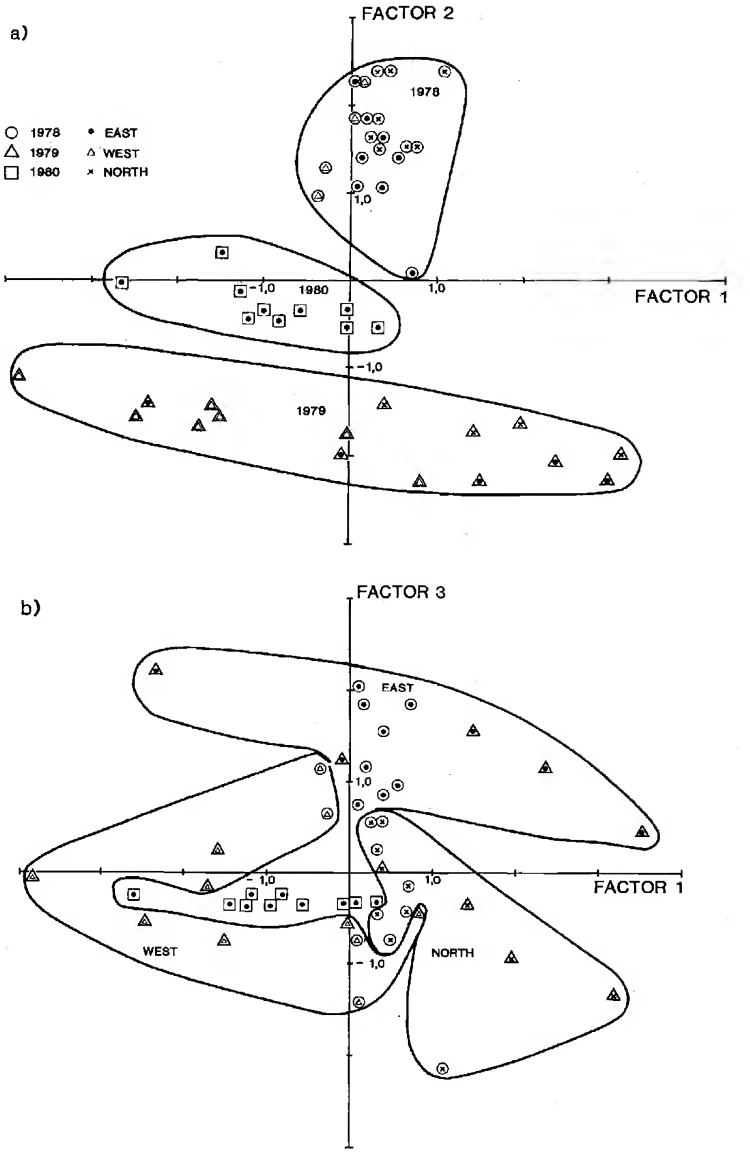


FIG. 13. PLS2 scores for the 46 swede samples analyzed as in Figs. 11 and 12. (a) Factors 1 and 2; (b) Factors 1 and 3.

chemical variables fructose (−), glucose (−), dry matter (−) and soluble solids (−). This variation was apparently independent of season, but weakly showing a variation between sites (i.e. west versus north). Factor 2 was dominated by the yearly variation in the material, describing the swedes from 1979 as more sulphurous and bitter than those from 1978. This may reflect a sensory flavour variation not easily measurable by the chemical methods. The third factor was obviously describing a sensory texture variation weakly related to the titratable acids (+) and pH (−). This reflected a variation especially in chewing resistance (+) and juiciness (−) between swedes from the east of Norway (+) and the north (−). The fourth factor was small and of minor interest in this connection. Further interpretation of these results is beyond the scope of the present paper.

A PLS-analysis without the design variables included in the **X**-block showed that only about 20% of the sensory variation covaried with the chemical variables, confirming factor 1 in the total PLS-analysis above as the main relationship. Similar effects by using design variables were observed in a corresponding study of carrots (Martens, M. *et al.*, 1985).

This PLS-analysis illustrates a general way of approaching the interpretation of PLS2 results: Although a small *average* variation in **Y** in the first factor (only about 4% as seen from Fig. 11) was explained, certain individual **y**-variables were better explained. For instance, about 23% of the variation in sweetness was explained in factor 1; after four factors about 43% was explained. However, using a PLS1 analysis to study y=sweetness more specifically improved its percentage explained variance to about 68% after its three significant factors.

Conclusions: PLS2 may be useful for getting a total picture of different relationships between many **Y**- and many **X**-variables. A weak relationship between sensory and chemical variables was found for swedes; only about 20% of the total variation in the sensory variables could be predicted by the chemical variables alone. A relatively strong variation between seasons and sites caused an increase of the percentage explained variance in the sensory **Y**-block to about 46%.

After getting a total overview, within-season and within-site variabilities need to be further investigated. The design variables were useful for interpretation of the loading plot, in a way replacing the scoring plot. Modelling a single **y**-variable by individual PLS1 regression on the same **X**-variables gave improved prediction ability.

4.4. PLS2 with Many Y-variables and Different Numbers of X-variables and Objects

Example: Explore how combinations of chemical and physical measurements improve prediction of sensory quality of cauliflowers (*Brassica oleracea* var. *botrytis* L.). Study the effect of having few objects in the training set.

Model: In the PLS2 in Fig. 4(b), model $J=12$ sensory variables (**Y**-block) from between $K=1$ and $K=8$, i.e. 1–8 chemical and physical variables (**X**-block), over $I=27$ or $I=6$ objects.

Experimental: The 12 sensory attributes (total colour, total appearance, crispness, chewing resistance, juiciness, total texture, sweet-, fruity-, bitter-, sulphurous flavour, flavour strength and total flavour) were evaluated for each sample in three replicates. The seven chemical measurements (dry matter, pH, total titratable acids, vitamin C (ascorbic acid), fructose, glucose and sucrose) were performed in two replicates and calculated as percentage of fresh weight. The physical variable, weight (g) of each sample, was taken as a mean of 15 individual cauliflower heads. Details about the sensory and chemical analyses are given elsewhere (Fjeldsenden *et al.*, 1981; Martens, M. *et al.*, 1983a,b).

The different samples were chosen in order to include variation due to season, sites and variety.

Results and discussion: Fig. 14 shows percentage explained variance in the **Y**-block for six PLS2-analyses, calculated by cross-validation, using various **X**-variables alone and in combinations. Only about 4% of the variation in the sensory variables could be explained by the **x**=weight alone. A slight increase was observed for (**X**=weight and dry matter), while the sugar analyses (**X**=sucrose, fructose and glucose) alone seemed to describe the same *amount* of variation. But these two groups of **X**-variables did not describe the same *type* of variation. This was seen by combining (**X**=weight, dry matter and the three sugars). While the weight and dry matter reflected sweetness, crispness and juiciness, the sugars were apparently related to the sweet–bitter balance. The 'acids' variables (**X**=total titratable acids, pH and vitamin C) were relatively strongly related to the sensory variables, first of all fruity flavour, but also crispness and juiciness were explained. However, the combination of all the chemical and physical variables were found to predict the total variation in **Y** optimally; about 75% variance explained after three factors.

The latter PLS2 analysis predicting the 12 sensory variables from all

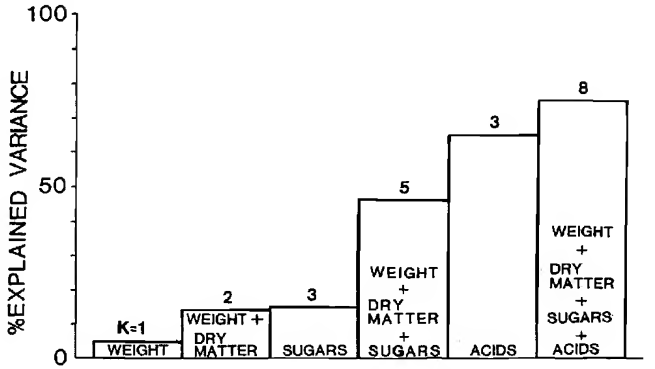


FIG. 14. Percentage explained variances of 12 sensory **Y**-attributes of cauliflower (27 samples) described by chemical and physical variables alone and in different combinations ($K=1$–8). Averaged results after 1–3 PLS factors found to be significant are shown.

the eight chemical/physical variables, was repeated on a lower number of samples: six selected samples were chosen to span the main variations. This PLS2 analysis gave PLS loading plots very similar to those obtained by analyzing all 27 samples (not shown here).

Conclusions: PLS2 regression allowed us to model **Y** by a low-number of 'harmonies' $\hat{\mathbf{T}}$ from an 'orchestra' of instruments **X**. Multicollinearities between the **X**-variables were no problem. Sensory quality of cauliflowers was found to be best described by a combination of dry matter, sugars and acids variables. PLS worked well even on small sample sets with fewer samples than variables.

4.5. PLS2 with only Design Variables in the X-block

Example: Explore relationships between demographic variables of consumers and their preference of different types of peas (*Pisum sativum* L.).

Model: In the PLS2 in Fig. 4(b), model the preference for $J=16$ pea batches (**Y**-block) from $K=11$ demographic design variables (**X**-block) over $I=96$ objects (i.e. 'samples'=consumers). For simplicity, design variables for interactions were not used. An opposite model with only design variables in the **Y**-block will be discussed in Section 5.5.

Experimental: The ability to predict the average consumer preference by a laboratory sensory panel was studied in Section 4.2. The individual response from the 96 consumers were now further investigated: in addition to the hedonic evaluations of the 16 pea batches, a questionnaire gave information about sex (2), age (4), frequency of use (3) and place of belonging (2) for each person; the numbers in parenthesis refer to the categories of each demographic variable (in total 11). These 11 **X**-variables were used for modelling the disagreements between the people on the different pea batches.

Results and discussion: On average about 20% of the between-people variation in the preference of the peas was accounted for by differences in sex, age, etc., among the consumers (two PLS-factors). Place, sex, age and frequency of use were described to 87%, 60%, 38% and 20% respectively by these two factors. Figure 15 shows that the **X**-variables age, place and sex were contributing the most to PLS factor 1; i.e. there was a tendency that young people (age 15–25), men and people from place 2 seemed to prefer peas of low maturity (low TV). Elderly people (age above 60), females and people from place 1 apparently accepted peas of higher maturity. In factor 2 a weak relationship was revealed between preference of certain peas by consumers that 'often use' peas compared

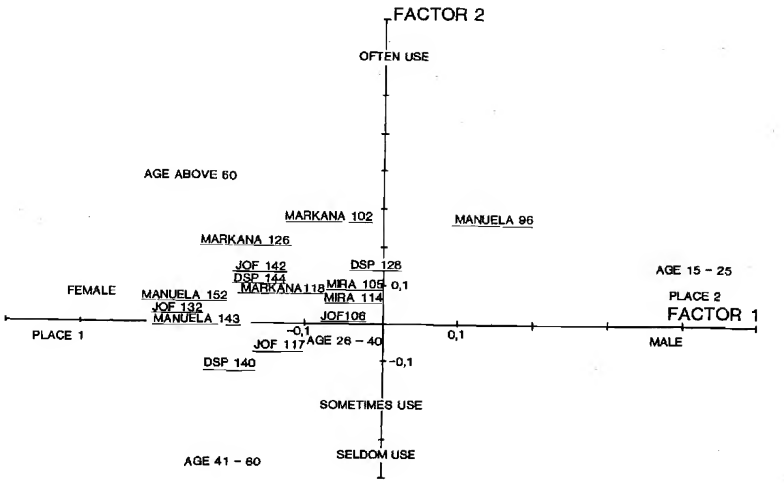


FIG. 15. PLS2 loadings for factors 1 and 2 for 16 preference variables on different pea samples (**Y**-block) and 11 demographic variables (**X**-block) registered for 96 consumers. The **Y**-variables are underlined and marked according to pea variety and maturity level (tenderometer value, TV).

to consumers that 'seldom use' peas, and people in the 'age 41–60' group. The fact that elderly people also seemed to contribute to preference of low maturity peas, confirm other statistical analyses indicating that consumers of that group were split according to frequency of use. This might have been revealed if interaction variables, e.g. age × usage frequency, had been included in the study. Further, Fig. 15 shows that consumers' preference for a certain pea variety was not evident from the present linear modelling, and that consumers of age 26–40 and people only sometimes using peas, showed preference pattern in between those of the other age and usage groups. Further PLS discussions with respect to the experimental design and potential consequences are left out here.

In summary, the results in Section 4.2 showed that a large proportion (73%) of the *between-batches* preference variations was systematic enough to be predicted linearly from laboratory sensory data; in that example the average of the consumer results were used as **y**, and the PLS1 showed that sensory quality appeared to be important for the consumers. In the present extension of the same example we studied the *between-people* preference variations for the different pea batches. Now we see how people disagree on the different pea qualities. Only 20% of their disagreement variance could be explained by their sociological variables. Additional information about people is probably required in order to predict their differences in preference. But one other reason for the low prediction ability is probably the fact that the 16 **Y**-variables were standardized; this amplifies the noise from pea batches on which people agreed. Furthermore, we have used a straightforward linear metric PLS model; improved modelling might be possible if software for non-metric PLS had been available, e.g. based on the Optimal Scaling principle (Young, 1981).

Conclusions: A weak but interpretable and valid relationship was found between people's preference disagreements concerning peas and demographic information about these people with respect to, for example, age and sex. A non-metric PLS may be relevant for similar problems.

4.6. Multi-block Consensus PLS (CPLS)

Example: Find the main common information in multi-block sensory data in order to correct for systematic response differences between judges on different sensory attributes.

Model: In the CPLS model (Wold, S. *et al.*, 1986) in Fig. 7(a) let $J=7$ judges, $K=6$ sensory texture variables and $I=5$ alginate solutions. How

can 7 **X**-blocks ($\mathbf{x}_j, j=1, 2, \ldots, 7$) efficiently be combined into a common concentrated consensus table with respect to I objects?

Experimental: The 6 sensory variables (1 = viscosity, 2 = hardness, 3 = cohesiveness, 4 = body, 5 = resistance to deformation, 6 = smoothness) were evaluated for 5 alginate solutions of different concentrations (4·0, 2·0, 1·3, 0·65 and 0·25% weight/volume). Data from 7 judges were used (6 judges A–F plus a seventh 'judge' G, G representing the average of 12 judges).

This experimental part was taken from a larger study on relationships between fundamental rheological data and sensory perceived mouthfeel, Bohlin *et al.* (1985), where further details about the experiment as well as results with focus on sensory–instrumental relationships by PLS2 are described.

Results and discussion: In the CPLS analysis (Fig. 7(a)) the average was first estimated and subtracted in order to compensate for systematic differences in the general level of results. Then the CPLS algorithm estimated a series of orthogonal consensus scores $\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \ldots,$ each with their associated people × terms rotation and scaling matrix, loading $\hat{\mathbf{P}}$, their samples × people individual score matrix $\hat{\mathbf{U}}$, and their individual consensus influence weights $\hat{\mathbf{q}}$. After each factor the residual variances were computed, with respect to the K terms, the J people and the I samples.

The leverage-corrected total residual variance revealed only one CPLS factor to be significant, explaining 73% of the variation in **X**. But for illustration, the next two factors are also discussed. Figure 16(a) shows that terms 1, 4, 5 and 6 were well explained in factor 1; there was drastic reduction in rest variances from 0 to 1 factor. Terms 2 and 3 had low initial variance; they were not important with respect to describing the texture variation in the material. Especially term 6, but also terms 4 and 5 were modelled by factor 2. Looking at the judges' contribution to each factor (Fig. 16(b)), all of them reflected the type of variation described in factor 1, but to different degrees. Factor 2 revealed some confusion among the judges, in particular, judges A and E, and to a smaller extent C and D, differed from the average (i.e. 'judge G').

These differences between the judges (A–G) may be understood from the CPLS loading plot for the first two factors (Fig. 17). Along $\hat{\mathbf{P}}_1$ all of the judges seemed to agree (circled results). They even seemed to span the 1–9 point scales rather similarly. It may be noticed that this main variation is also reflected in the average value (G). Term 1 contributed strongly and positively in a consistent way (judges A–G grouped to-

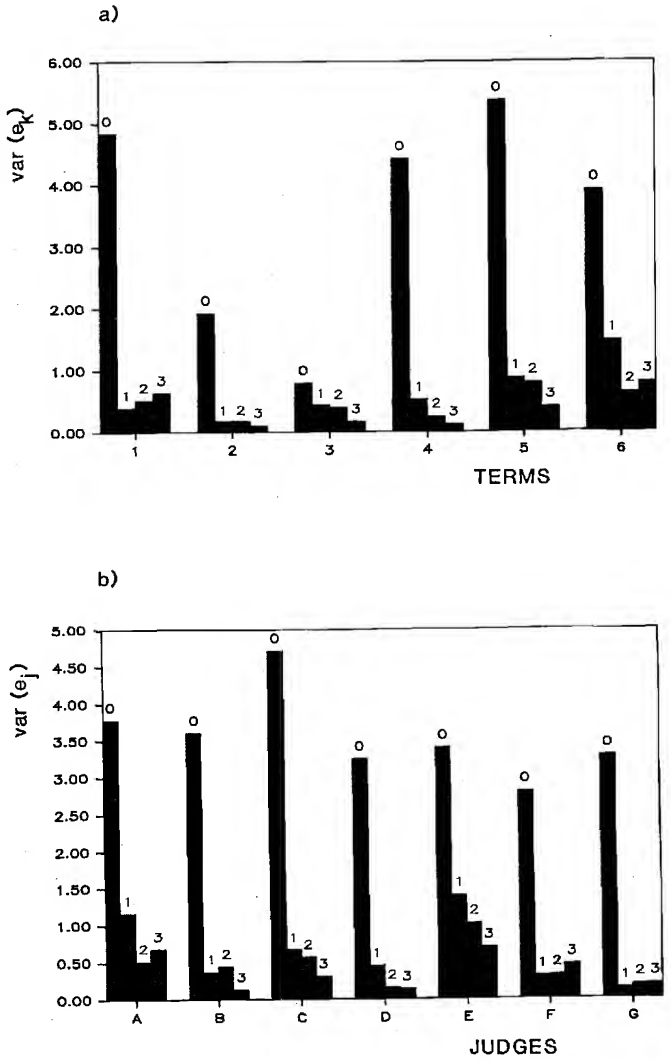


FIG. 16. Rest variances in $\hat{\mathbf{E}}$ after 0, 1, 2 and 3 CPLS factors: (a) for six different sensory terms (1–6); (b) for seven different judges (A–G).

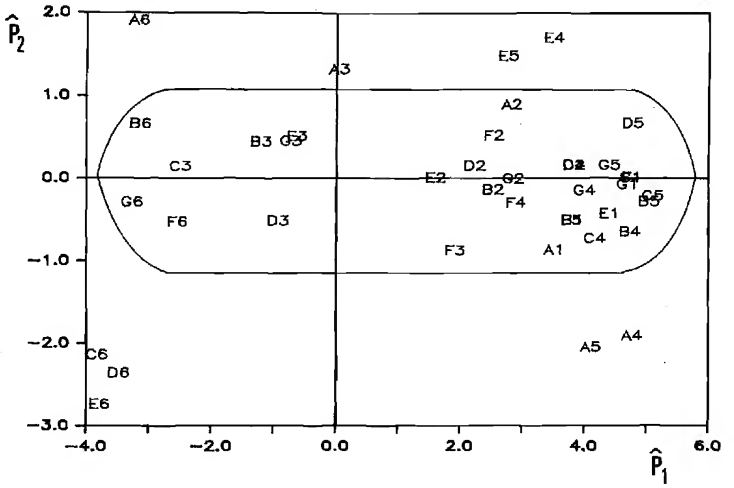


FIG. 17. CPLS loadings $\hat{\mathbf{P}}$ for factors 1 and 2. The letters (A–G) refer to different judges and the numbers (1–6) to different sensory terms.

gether), term 6 contributed strongly and negatively, while terms 2 and 3 didn't vary much with this consensus factor. In factor 2 judges A and E used the terms 4, 5 and 6 in an opposite way; for example, when A scored higher than the others on term 6, then E scored lower and vice versa. Term 6 (smoothness) in general was perceived differently by the judges. Factor 2, being minor and non-significant, should not be paid attention to when interpreting the results about the alginate solutions.

The results obtained were easy to interpret by plotting the CPLS factor 1 scores versus the alginate concentrations. (This corresponds to relating $\hat{\mathbf{t}}_1$ to $\mathbf{Z}$ in Fig. 7(b).) Figure 18 shows the consensus $\hat{\mathbf{t}}_1$ to be a nice, smooth and non-linear function of alginate addition: with increased alginate concentration the solutions became first of all more viscous (term 1) and less smooth (term 6), but also a tendency toward stronger body (term 4) and higher resistance to deformation (term 5) was evident.

To study the validity of this first CPLS solution further, a general evaluation of the judges along a good–bad scale (10 = very good; 0 = no good) was done by sensory staff not involved in the data analysis. These results, corresponding to the $\mathbf{Y}$-block in Fig. 7(b), are plotted versus the

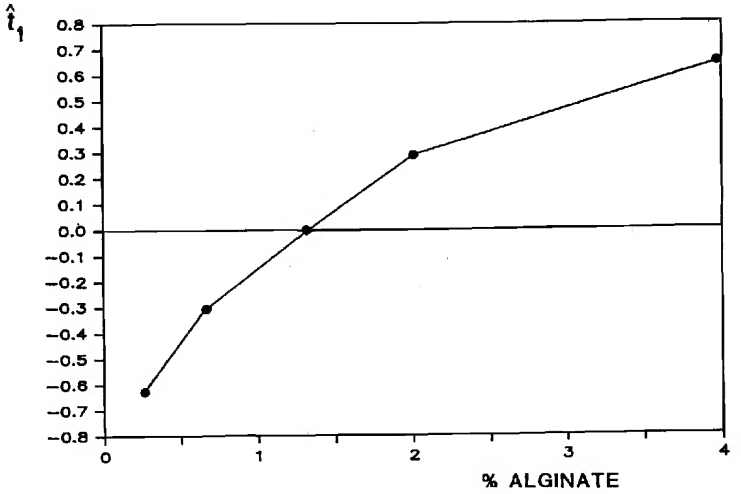


FIG. 18. CPLS scores $\hat{t}_1$ for factor 1 plotted against alginate concentration in the five different samples.

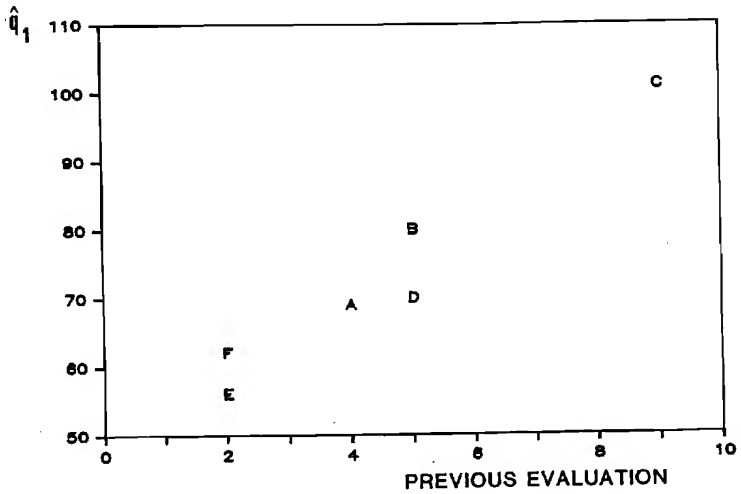


FIG. 19. Weights $\hat{q}_1$ for the CPLS factor score 1 plotted against an independent previous evaluation of judges A–F.

weights for the CPLS factor score 1, $\hat{q}_1$ (Fig. 19). This shows that the best judge in the panel also contributed the most to the consensus solution $\hat{t}_1$.

Since this was a small data set with a relatively simple structure, CPLS needs to be tested on other data sets as well before its real usefulness is clear. Also, the CPLS should be compared to Procrustes analysis (see Chapter 7 in this book).

Conclusions: CPLS was found to extract and concentrate the essential information from a multi-block system, thereby 'cleaning' up noisy data before relating them to, for example, instrumental data. CPLS seems to be a promising method to study the performance of judges in a sensory panel. It may also have relevance in comparison of different analytical techniques, e.g. multivariate ring tests.

5. APPLICATIONS OF PLS VERSUS OTHER METHODS

5.1. PLS versus Multiple Linear Regression (MLR)

Principles: For modelling quantitative relationships between a **y**-variable and a block of **X**-variables, the conventional statistical method is MLR. The distinction between MLR and PLS is described in Section 2.1. Because of a mathematical peculiarity MLR cannot use highly intercorrelated **X**-variables, therefore a number of different methods have been developed for eliminating apparently redundant **X**-variables (Stepwise Multiple Linear Regression, SMLR).

These SMLR methods are routinely used in standard statistical software packages (often without proper validation routines). It is often difficult to choose which and how many **X**-variables to use, and even more difficult to interpret the result, since we then do not know the importance of the **X**-variables eliminated. PLS regression offers an easier and better alternative, as the next example will show.

Example: We want to predict the degree of sweet taste (**y**) in cauliflower from a set of seven different chemical variables $\mathbf{x}_1, \ldots, \mathbf{x}_7$. The sweetness was measured by a trained panel of 12 judges in three replicates and averaged, while the seven chemical variables were measured by various conventional methods, averaging two replicates.

Data were available for 12 well selected cauliflower samples. A conventional upward-selection SMLR program was first used. Different criteria in the program indicated different dimensionalities of the optimal solution, but a four-dimensional regression appeared to be the most

sensible (F-test was significant). The four **X**-variables selected were total titratable acids (x_1), vitamin C (x_2), glucose (x_3) and sucrose (x_4).

But was this four-dimensional calibration model really optimal for predicting the sweetness in future cauliflowers of the same type calibrated for? Figure 20 shows the leverage-corrected prediction error as a function of the dimensionality of the PLS model based on the four **X**-variables selected by the SMLR program. The 'noise line' at variance 0·03 indicates the expected uncertainty of the sweetness data themselves, based on between-replicate variances.

The SMLR program indicated an estimated error variance of 0·03. This four-regressor MLR solution, i.e. the present SMLR, is mathematically identical to the full four-factor PLS solution ($\hat{\mathbf{E}}=0$). The validation showed that even this SMLR gave overfitting of the data. The SMLR prediction error seen from Fig. 20 (same as prediction error after four PLS-factors) turned out to be 0·08, almost three times higher than what the original SMLR program had indicated! The PLS program, in contrast, showed that a two-factor solution had better predictive ability, variance = 0·05, corresponding to ±0·22 on the 1–9 point sweetness scale.

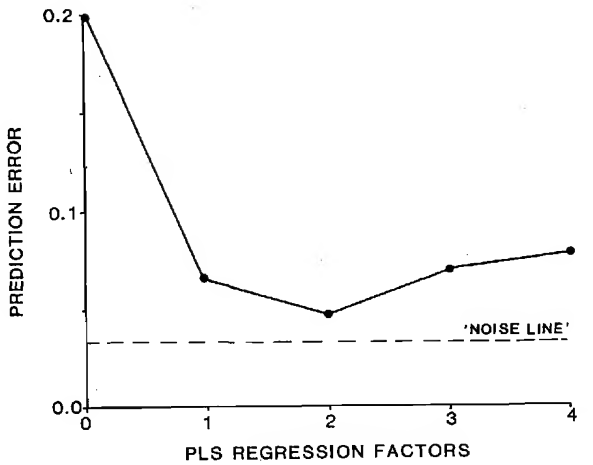


FIG. 20. Prediction error for sensory sweetness (y) of cauliflower, given as mean-square error as a function of the number of PLS regression factors. 'Noise line' represents the measurement noise (standard error of the mean for the three sensory replicates).

The prediction error is a function of $(\mathbf{y}-\hat{\mathbf{y}})$ (see Section 6.2). The dotted 'noise line' at variance 0·03 gives the expected uncertainty of the input sweetness assessments **y**. Below this line the prediction error cannot ever be expected to come, even if the number of samples in the training set had been very high. Actually, the prediction error must be higher than this y-noise alone, because it also contains more or less errors from $\hat{\mathbf{y}}$, e.g. due to non-linearities and to errors in the estimations of the model parameters, as well as prediction errors due to noise in **X**. Thus it was not surprising to find that the apparent SMLR variance 0·03 (similar to the 'noise line'), was too optimistic due to overfitting. In the above PLS1 regression the four **X**-variables were used unscaled.

Figure 21 shows the estimated coefficients $\hat{\mathbf{b}}$ from the two-factor PLS solution, compared to the original SMLR coefficient estimates. The two estimates have the same general pattern, opposing the sucrose against the other three variables. But the SMLR-based coefficients are much larger than the PLS based ones, indicating that the SMLR, even after the

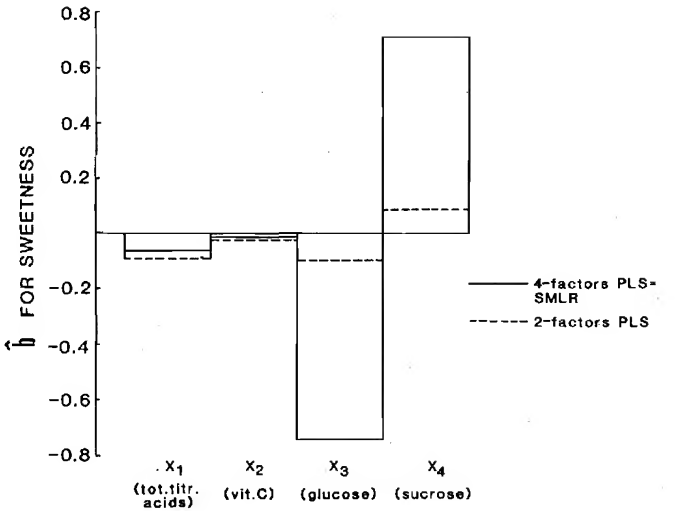


FIG. 21. Estimated regression coefficients $\hat{\mathbf{b}}$ for predicting sensory sweetness from the four chemical **X**-variables, given for the four-factors PLS solution, which is equal to the SMLR solution (solid lines), and the optimal two-factors PLS solution (dotted lines).

stepwise elimination procedure, overemphasizes minor variations in the data and thereby overamplifies noise and irrelevant phenomena.

Conclusions: In this example we first used conventional SMLR to select the 'best' subset of **X**-variables. Then we applied PLS1 regression on these to evaluate this solution. The results indicated that the SMLR had been overfitted. The example shows that PLS1 regression may be easier to use than SMLR. It also illustrates the importance of validating the results with respect to prediction error. This is seldom done in regression software.

5.2. PLS versus Canonical Correlation Analysis (CCA)

Principles: CCA is a statistically oriented two-block method analogous to PLS2; it constructs linear combinations of the variables from the **X**-block that correlate maximally with the **Y**-block and vice versa. The computation requires that the number of objects is higher than the number of variables in **X** and **Y**. In essence both PLS2 and CCA estimate a small number of factors or dimensions in order to express the systematic variations common to the two data matrices, the **X**-block and the **Y**-block in Fig. 22. The loadings of the variables in each factor are correlations (CCA) or modified correlations (PLS).

The major difference between the two techniques concerns the way these factors are estimated from the raw data, as popularly illustrated in Fig. 22. If there are many variables in **X** or **Y** compared to the number of objects, then CCA first requires the use of a rank reduction step (PCA, factor analysis, etc.) prior to the CCA analysis. In PLS regression these two steps are done simultaneously in one algorithm. This ensures statistical parsimony.

Hence, from a statistical point of view the two techniques are expected to respond differently to multicollinearities and to the relative number of variables compared to objects. One other difference, indicated by the horizontal arrows in Fig. 22, is that CCA is a purely correlative method, while PLS2 regression gives a predictive direction from **X** to **Y**. The PLS algorithm can be modified to a correlative one if desired.

Example (from Martens, M. and Burg, 1985): The following data set was studied by PLS2 regression and CCA: 34 pea samples were analyzed by 12 sensory variables (**Y**-block) and 14 chemical and physical variables (**X**-block). Thus, $I=34$, $J=12$ and $K=14$ variables.

PLS2 regression revealed four significant factors with which about 74% of the total variation in the **Y**-block was explained by the **X**-block (Fig. 23(a)).

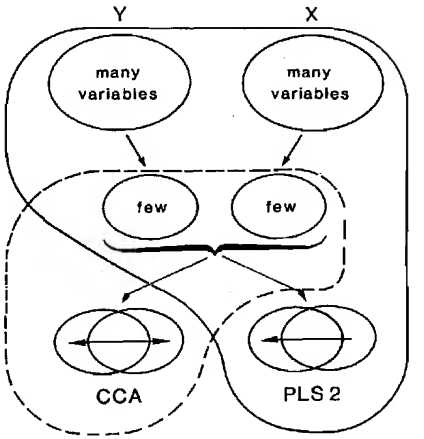


FIG. 22. Conceptual difference between CCA and PLS2 regression. If the number of variables is larger than the number of objects, then CCA needs a prior variables reduction step (e.g. PCA, factor analysis) before relating the two blocks of variables, while PLS2 is doing this simultaneously. The PLS regression has a predictive direction while CCA is correlative.

A non-metric CCA used directly on the data, failed due to rank problems; the number of objects was not high enough. A non-metric PCA (PRINCALS) was therefore first used on each of the blocks separately to reduce the number of variables. This gave two intermediate score matrices, each describing a few principal factors from **X** and **Y**, respectively. These were subsequently related to each other by a metric CCA, which revealed two significant canonical dimensions. With this two-factor solution about 67% of the total variation in the **Y**-block was explained by the **X**-block (Fig. 23b).

Summarized from Fig. 23(a, b) the PLS and the PRINCALS-CCA solution gave mainly the same results and were, as such, both suited for studying relations between two blocks of data. However, this example as well as other references stress the necessity of keeping the number of variables low in CCA, as in MLR. This is a limitation of serious consequences since modern analytical instruments very easily and relatively cheaply give data on a large number of variables on each object. But getting many objects, especially for analysing food samples or biological material in general, may be difficult and expensive. This

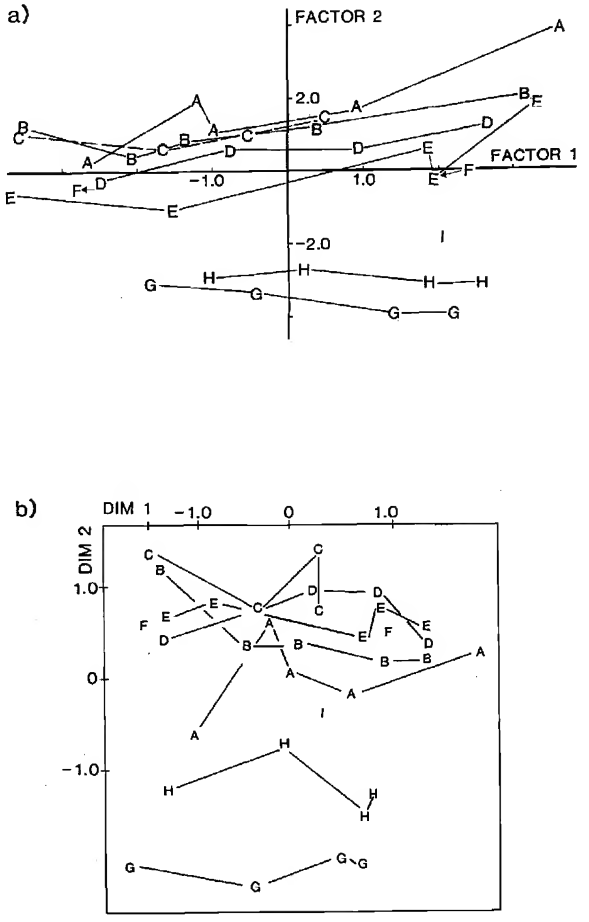


FIG. 23. Results from 12 sensory variables (**Y**-block) and 14 chemical/physical variables (**X**-block) measured on 34 pea batches, marked according to varietal name in the figures: (a) PLS scores for factor 1 and 2; (b) CCA mean canonical scores for dimensions 1 and 2. Factor or dimension 1 describes maturation of peas (solid lines connecting the same pea varieties) from right to left, while factor or dimension 2 describes differences between varieties. (Reprints from Martens and Burg, 1985, with kind permission from Elsevier, Amsterdam).

problem was exemplified by Canaway *et al.* (1984) in an attempt to use CCA to establish relationships between individual vocabularies from sensory profiling. They apparently had to carry out the canonical correlation analysis on the replicates in order to get enough 'samples', and thereby obtained results that were difficult to interpret. For a similar problem connected to sensory profiling, Martens, M. and Russwurm (1980) used factor analysis to reduce the number of variables below the numbers of samples prior to relating the sensory to chemical data by CCA. Piggott and Jardine (1979) used CCA successfully to relate sensory descriptive terms of whisky to chemical constituents, and CCA gave interpretable results in a study of odour perception of food-relevant chemical components (Schiffman, 1981). In an attempt to relate dynamical criteria from a baking process to baking quality variables, Möttönen (1975) concludes that CCA was not powerful for studying quality.

5.3. PLS versus Multidimensional Scaling (MDS)

Principles: MDS is designed for the analysis of square tables where the number of rows equals the number of columns, for instance symmetrical similarity or dissimilarity tables. Chapter 8 in this book describes MDS; we shall here show that even PLS2 can be used for square tables.

Example: 132 persons from the food industry were asked to evaluate similarities between 20 flavour attributes on a 1–9 point scale, resulting in 132 symmetric 20×20 similarity matrices (Fig. 24).

A non-metric MDS (Young and Lewyckyj, 1979) was performed on the one-mode, two-way matrix representing the 20×20 averages over all 132 persons; treating these similarities as ordinal data.

The metric PLS2 modelling was performed by regarding the 20×20 average of the first 66 persons as **X** and the 20×20 average of the last 66 persons as **Y**. These **X** and **Y** similarity data (scale 1–9) were transformed to dissimilarities (scale 8–0). In order to reduce the effect of giving 'non-related' and 'opposite' the same high scores, the square roots of the dissimilarities were used in the PLS-analysis. Since the **X** and **Y** matrices are symmetric, they were double centred prior to the PLS-analysis, i.e.

$$e_{ik} = x_{ik} - \bar{x}_i - \bar{x}_k + \bar{\bar{x}}, \text{ where } \bar{\bar{x}} \text{ is the average of vector } \bar{\mathbf{x}}.$$

Six PLS factors appeared to be significant, judging from the similarity between $\hat{\mathbf{p}}$ and $\hat{\mathbf{q}}$. The scores of the 20 attributes along the first two factors are shown in Fig. 25(a). Figure 25(b) correspondingly shows the first two factors or 'dimensions' from a six-factor MDS solution. In both plots factor 1 separates the fruity, aromatic quality properties from

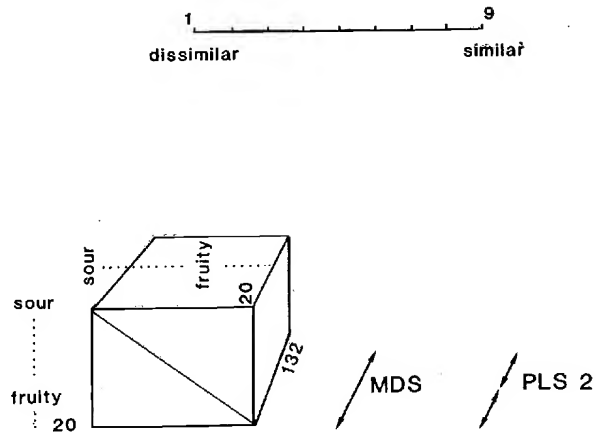


FIG. 24. Conceptual difference between MDS as a one-block method, and two-block PLS2. Experimental design of a study of similarities between 20 different flavour attributes evaluated by 132 persons using a 1-9 point scale.

rancid, metallic, etc., while factor 2 separates sour, acid from boiled, oily flavour attributes.

After six factors the non-metric MDS modelled 92% and the metric PLS modelled only 73% of the total variation in the perception data. (A corresponding six-factor PLS solution for the same data without having taken square root, explained only 56%.)

A combination of MDS and PLS as a link between, for example, a cognitive perception matrix *and* some external demographic variables or chemical and physical variables, opens up many possibilities. Works on such PLS combinations are in progress. Hoffman and Young (1983) successfully used a non-metric MLR to interpret MDS-results from a study of different persons' perception of various beverages.

The ALSCAL program can also analyze three-way two-mode square matrices, using the 'INDSCAL' MDS model. The three-way CPLS algorithm can probably be used in a similar way, replacing '*I* objects' in Fig. 7(a) by '*K* terms'. The scores $\hat{\mathbf{T}}$ would then show the consensus between the *J* people, concerning the relationships between the *K* terms.

Conclusions: The MDS explained much more variance than the PLS regression, probably because the former allowed for non-metric characteristics in the input data. But the plots of their main parameters were

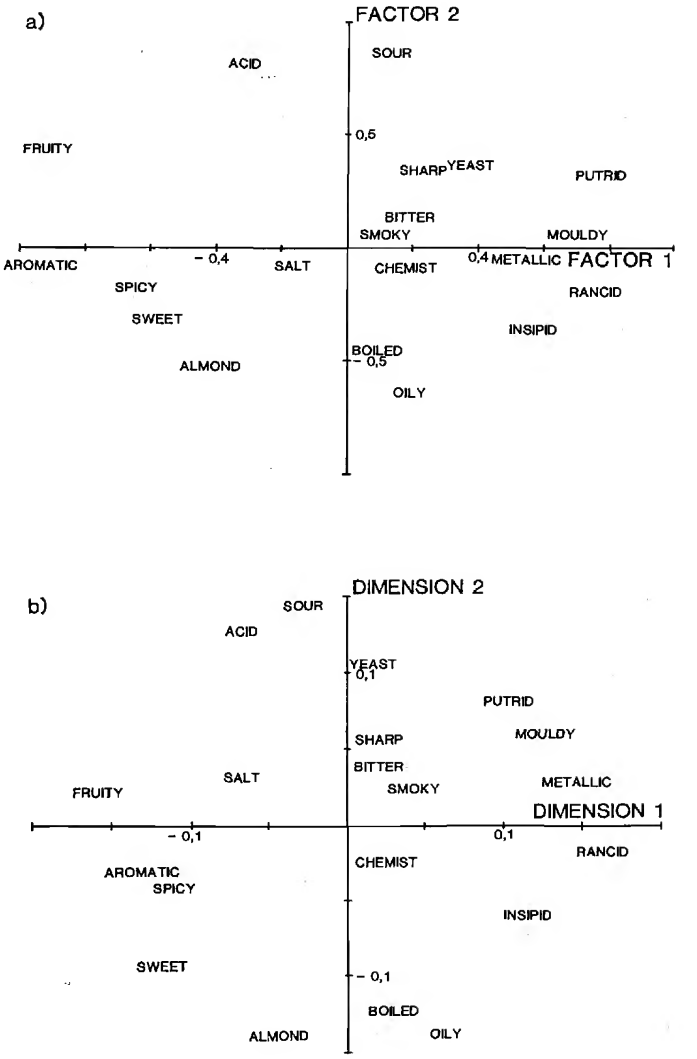


FIG. 25. Results from the study described in Fig. 24: (a) PLS scores for factors 1 and 2; (b) MDS coordinates for dimensions 1 and 2.

quite similar. This indicates that although a non-metric PLS program is desirable, even the metric PLS regression can give meaningful modelling of non-metric data. It also indicates that PLS is useful for relating a symmetric matrix to another symmetric (or non-symmetric) matrix.

5.4. PLS versus Analysis of Variance (ANOVA)

Principles: In ANOVA, a response variable, **y**, is regressed on a set of design variables, **X**. The PLS regression here maps the salient features in the design matrix in order to provide a visual representation of the ANOVA significant tests.

Example: Factors influencing preference of black currant juice were studied and the results analyzed by ANOVA (Martens, M. *et al.*, 1983c). 32 samples were subjected to a factorial design of five main factors, each in two levels: *Colorant* (regular-added green colorant), *acidity* (regular-added citric acid), *volume* (15–25 ml), *temperature* (+5 °C–+20 °C) and *time* of serving (before–after lunch, i.e. AM–PM). 24 consumers evaluated their degree of liking on a 1–9 point hedonic scale.

A five-way ANOVA showed significant differences between three of the main factors (Table 1). For the PLS2 regression each of the five factors in two levels was described by two design variables in an **X**-block (e.g. regular colour: $x_1 = 1$, $x_2 = 0$; added green colorant: $x_1 = 0$, $x_2 = 1$). Thus, $I = 32$, $K = 10$ and $J = 24$. The results from Table 1 could then be directly seen in the PLS2 loading plot (Fig. 26(a)). The two first PLS2 factors (25% and 7% explained cross-validated variance in **Y** respectively) mapped the variation in colorant (PLS factor 1) and acidity (PLS factor 2), the two most significant experimental factors from the ANOVA. The individual consumer's preference may also be studied from this loading plot, e.g. person No. 14 particularly disliked black currant juice with green colour but liked it somewhat sour. Persons No. 8 and 24 had

TABLE 1

MAIN EFFECTS FROM A 5-WAY ANOVA FOR BLACK CURRANT JUICE; 32 SAMPLES (2^5 FACTORIAL DESIGN) EVALUATED BY 24 PERSONS

Factor	*Order of mean value*	*Level of significance*
Colorant	Reg. C. > Green	$p < 0{\cdot}001$
Acidity	Reg. A. > Citric	$p < 0{\cdot}05$
Volume	25 ml > 15 ml	$p < 0{\cdot}05$
Temperature	5 °C ≃ 20 °C	not sign.
Time	AM ≃ PM	not sign.

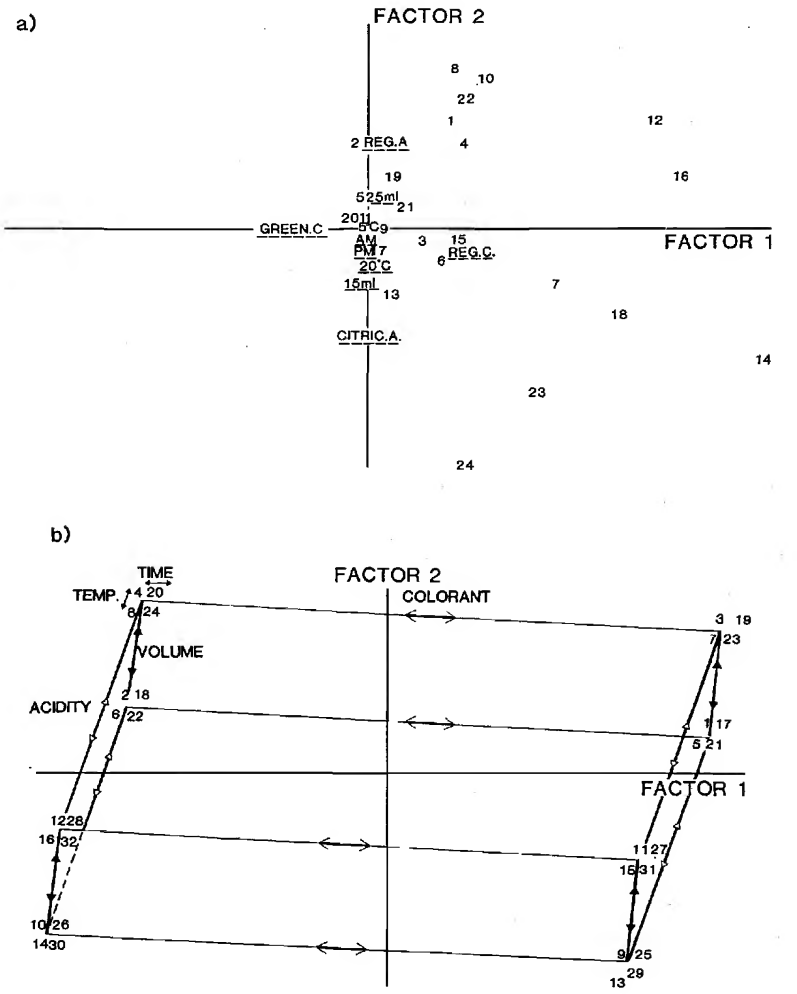


FIG. 26. Results from a study of relationship between five different design variables (X-block) and 24 persons' hedonic scores (Y-block) on 32 samples of black currant juice: (a) PLS loadings for the design variables (underlined) and the persons (1–24) for factors 1 and 2; (b) PLS scores for the samples (1–32) for factors 1 and 2.

opposite preferences concerning acidity. The scoring plot (Fig. 26(b)) nicely confirmed that preference of black currant juice was found to be mainly influenced by colorant, then acidity, but also an effect of volume was significant.

For simplicity the interactions in the ANOVA are not discussed here. However, including the interactions as design variables in the **X**-block would probably have made the interpretations even more detailed.

5.5. PLS versus Classification Techniques

Principles: By projection of the **X**-variables down on a set of design (or dummy) **Y**-variables, PLS may be used as a classification method such as linear discriminant analysis (LDA), canonical variate analysis (CVA) and others. PLS classification can be done in two different ways. (1) MACUP, the PLS equivalent of the PCA-based SIMCA classification, relying on **X**–**Y** relations instead of only intra-**X** relations. (2) Discriminant PLS regression, using design variables **Y** to represent classification. In the SIMCA and MACUP classification methods (Wold, S. *et al.*, 1983b, 1984), each class is separately modelled by a principal component analysis or PLS regression, respectively. Classification is then obtained by calculations of distances from an object to every class model. The discriminant PLS regression tells us to which degree each of the variables contributes to a certain class separation, i.e. the loading plot gives the discrimination power of the **X**-variables (see Wold, S. *et al.*, 1983b, 1984). When expressed relative to the noise in the variables, the $\hat{\mathbf{b}}$-vectors summarize the discriminatory power of the variables.

In contrast to LDA, the PLS discriminant technique works even if the number of variables is large compared to the number of objects. While in LDA and PLS discriminant analysis the classes are assumed to be linearly separated, the MACUP has no restrictions as to asymmetric data structure.

Examples: The SIMCA classification method has been compared to LDA, CVA and other similar techniques with respect to food data (Forina *et al.*, 1983; MacFie, 1983) while the PLS-based classification techniques have not been applied much by food scientists. PLS outlier detection techniques, representing simple MACUP classifications, have been developed for calibration (Wold, S. *et al.*, 1983b; Martens, H. and Næs, 1986) and applied to food-relevant spectra: Martens, H. and Jensen (1983) distinguished between wheat and barley by NIR reflectance spectra; Martens, H. *et al.* (1986) revealed heme protein abnormalities in uv/vis transmission spectra. Discriminant PLS analysis (Wold, S. *et al.*, 1983b) was applied by Martens, M. (1986a) in a study of sensory and chemical quality criteria for stored versus not-stored frozen peas by treating 11 sensory and 9 chemical variables as the **X**-block and using design variables (1 and 0 for stored and not-stored, respectively) as the **Y**-block.

6. VALIDATION WITH RESPECT TO PRACTICAL INTERPRETATION

6.1. Optimal Model?

For a given set of data and a given application purpose it is important to check that a suitable mathematical model has been used. The 'soft' bilinear additive PLS model consisting of a series of orthogonal factors is very general. If several **Y**-variables are to be modelled, then a simultaneous PLS2 gives an overview, while a separate PLS1 solution for each individual **y**-variable often gives better prediction ability (see Section 4.3).

However, some sort of data pretreatment may be necessary to give an optimal fit of the data to the model; plots of parameters and residuals from preliminary PLS-fits to raw data will often indicate what type of pretreatment is needed. PLS regression may require some sort of weighting of the variables within a block to balance their signal/noise ratios. If the actual error level of a variable $\mathbf{x}_k$ or $\mathbf{y}_j$ is very different from the error level assumed in its weighting, this will decrease the prediction ability of the PLS results. But such error can automatically be detected by checking the estimated residual variances of the individual variables after the PLS-fit. Similarly, outliers among the samples $i=1, 2, \ldots, I$ can automatically be detected after the PLS-fit, in part based on their residual variance in **X** or **Y**, in part based on their leverage h_i (see Section 3.6). Abnormal individual data elements x_{ik} or y_{ij} can likewise be detected. Once an outlier has been detected in the data, the user must decide whether to knock it out from the data set and repeat the PLS modelling, to obtain new data from this object or just to accept it as an extreme, but not erroneous data point.

Other types of pretreatment like CPLS or Procrustes rotation techniques appear to be useful for cleaning up sensory data before putting them into a two-block study (see Section 4.6). If there are purely multiplicative effects in the model, these ought to be removed in advance (Martens, H. *et al.*, 1983a; Geladi *et al.*, 1985) or made additive by logarithmic pretreatments. Other linearizations of the input data can also

sometimes improve the fit (see Section 5.3), although this is not so important for higher-precision data, for example, from diffuse fluorescence (Jensen *et al.*, 1982) or Near Infrared Reflectance (Martens, H. and Næs, 1986).

In general, it is important to note that data pretreatment should not be used to make nonsense data fit a model.

6.2. Optimal Modelling Complexity?

Data modelling of **Y** (or a single **y**) from **X** can be done at different levels of ambitions, ranging from general causal modelling via predictive modelling to *ad hoc* fitting (Aakesson *et al.*, 1974). In this chapter we emphasize predictive modelling and try to avoid *ad hoc* modelling but leave the causal explanation to the subsequent interpretation of the results. Optimal predictive modelling means achieving a minimum prediction error.

The prediction error ($\mathbf{y}$-$\hat{\mathbf{y}}$) contains several errors; model errors due to incomplete or invalid modelling, and estimation errors due to insufficient training set data and propagation of noise from the **X**-data, plus, of course, errors in the 'true' test-data **y** (see Fig. 27(a)).

As illustrated in two of the examples in this chapter (Figs. 11 and 20), the prediction error usually reaches a minimum after a certain number of PLS factors, i.e. a certain complexity of the calibration model. This curved behaviour of the prediction error is a general phenomenon, and is explained conceptually in Fig. 27(a): as more 'harmonies' $\hat{\mathbf{t}}_1, \hat{\mathbf{t}}_2, \ldots$ (PLS factors) from the **X**-data are extracted and used as regressors for **Y**, the model errors and hence the prediction error goes down. But at the same time the effect of measurement noise and other irrelevant phenomena in the **X** and **Y** data cause an increased statistical uncertainty in the predictions. The total prediction error therefore reaches a minimum when the advantage of modelling yet another phenomenon equals the disadvantage of becoming yet more sensitive to noise, etc.

The PLS modelling requires a satisfactory design of the training set, both for valid interpretation of the obtained parameters and for optimal prediction. Figure 27(b) illustrates conceptually how, for a given data-analytic problem, different qualities of the training set data give different prediction errors.

The prediction ability of the PLS1 and PLS2 methods shown in this chapter is optimal for new samples similar to the average, $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, and it may decrease strongly for samples outside the ranges of the training set. Thus, the prediction testing requires a sensible selection of test samples in

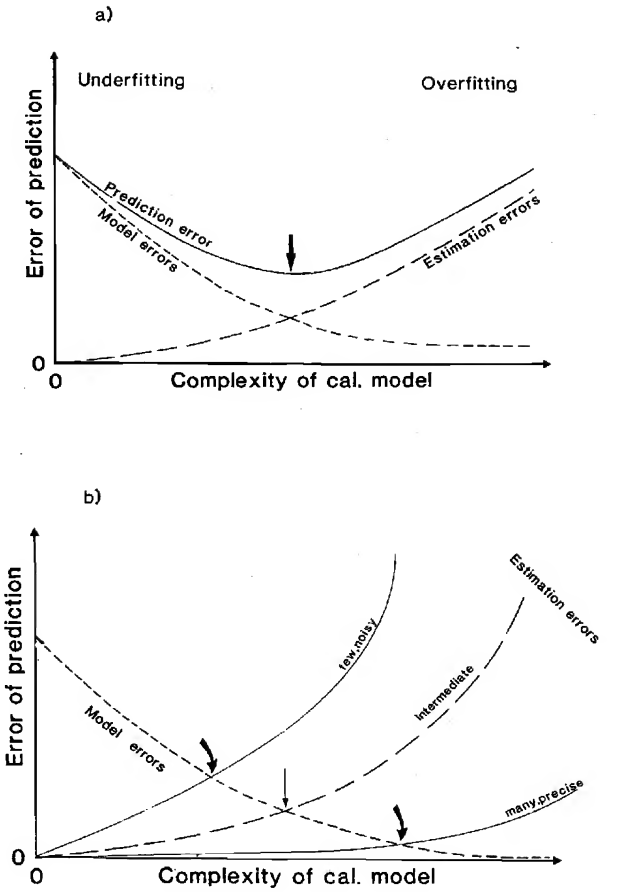


FIG. 27. (a) Conceptual illustration showing the total prediction error to be the sum of two main contributions: The model errors to be removed and the statistical uncertainty errors that can lead to overfitting. The arrow marks the optimal complexity, yielding minimal prediction error. (b) Effect of the quality and quantity of the input data: For calibration samples of a given type the optimal complexity of the solution (arrows) and the corresponding prediction error depends on the quality and quantity of data. Having few or noisy data gives worse prediction than having many or precise data.

order to give relevant results. In the case of cross-validation and of leverage-corrected prediction testing (see Section 3.6) the training set and the validation set are the same, and this puts extra requirements on the sampling: irrelevant, but systematic errors in the data will appear as valid factors. Therefore, the smaller examples in the present chapter (e.g. Section 4.6) must be regarded as method illustrations rather than conclusive experiments.

In general, the scientist must be prepared to accept zero significant factors as a possible outcome of a validation; in some data sets there simply is no valid information.

6.3. Practical Use of PLS Modelling

In addition to the mathematically based validation criteria (cross-validation and leverage-corrected variance), the data-analytic results must also be evaluated by other criteria: if outliers have been detected, should they be eliminated and the data reanalyzed? Are the statistically 'significant' results also practically significant and interpretable? Do they correspond to previous knowledge about the samples, the precision of the measurements, etc.?

In the practical interpretation of a PLS solution the user must take care to avoid scaling delusions due to data pretreatment. If the individual variables have been weighted (pre-multiplied by different coefficients), all the parameters $\bar{\mathbf{x}}$, $\bar{\mathbf{y}}$, $\hat{\mathbf{W}}$, $\hat{\mathbf{P}}$, $\hat{\mathbf{Q}}$ and $\hat{\mathbf{B}}$ will be affected accordingly, but can be de-weighted if so desired.

To ensure correct understanding, it can be useful to trace the obtained factor structure back to the raw data, as illustrated in Section 4.2 (Fig. 10).

7. PLS USED IN VARIOUS FIELDS

The PLS method, as described in this chapter, has recently been applied to other chemical and instrumental data with food research relevance, for instance:

* In physical chemistry: Predicting protein molecular conformation from circular-dichroism spectra; predicting food gel properties from fundamental rheological measurements; predicting mineral composition of rock samples from X-ray diffraction spectra; predicting particle size in powders from reflectance spectra.

* For 'classical' chemical data: Predicting concentrations of dyes from conventional UV/visual-range transmission spectra; predicting chemical composition of foods from NIR spectroscopy; predicting composition of protein mixtures from amino acid chromatograms (ion exchange and HPLC).

* In biologically oriented chemistry: Predicting chemical composition of living animals from X-ray computer tomography images; predicting botanical components in wheat flour mill streams from autofluorescence spectra; predicting growth of cereal varieties from climatic and agronomical information; predicting biochemical compositions from NIR spectroscopy in biotechnological fermentation.

PLS applications from the chemometric field outside food research are increasing rapidly (see Kowalski, 1984). Also, PLS has been found useful on ecological, economical, geological and medical data as well as data from the oil industry.

8. PLS PROGRAMS AVAILABLE

Since PLS is a fairly new multivariate data-analytic technique, it is not yet found in standard statistical packages. The PLS1 and PLS2 algorithms described here are implemented in the following two programs:

* The SIMCA-3B (and 3F) package developed by S. Wold, Umeå, Sweden (Wold, S., 1985). It is written in Microsoft Basic and runs on MS-DOS, PC-DOS and CP/M micros. The program is available from SEPANOVA AB, Östrandsvägen 14, S-12243 ENSKEDE, Sweden (Europe) and PRINCIPAL DATA COMPONENTS, 2505 Shepard Blvd, Colombia, MO 65201, USA. Fortran versions for minis and mainframes are available from SGAB, Box 801, S-951 28 Luleå, Sweden (PDP-11, VAX, CD Cyber, IBM, Prime, ...) and from O. Kvalheim, Chemical Institute, University of Bergen, N-5000 Bergen, Norway (VAX, NORD).

* The UNSCRAMBLER program package for multivariate calibration, used in most of the present examples, is being developed by H. Martens. It is written in Fortran 77 and runs on IBM–PC and VAX computers. The program package is available from Computer-Aided Modelling (CAMO A/S), P.O. Box 2893, N-7001 Trondheim, Norway.

While SIMCA-3B mostly aims at solving classification problems, the UNSCRAMBLER directs more towards calibration type of problems. Both programs have the characteristics listed in Fig. 28.

PLS is a desired method because it:

— handles many X-variables and Y-variables
— takes small sample sets
— takes missing values
— is realistic with respect to
  random noise in X and in Y
  multicollinearity in X and in Y
— gives warnings of outliers
— gives estimated parameters: easy to interpret
— gives predictive validation
— gives automatic determination of
  optimal model complexity

FIG. 28. Relating two blocks of variables: list of desired characteristics of a data analytic method from a user's point of view.

9. SUMMARY AND CONCLUSIONS

The predictive PLS1 and PLS2 regression methods and the new Consensus PLS analysis have been demonstrated on various food research data, especially for studies of sensory–chemical relationships. References to other applications are given. Advantages of PLS to more traditional statistical methods are exemplified. Validation with respect to practical interpretation is discussed.

Systems analysis based on the principle of Partial Least Squares regression offers new methods for studying relationships between two or more tables of variables measured on a set of objects. The methods avoid multicollinearity problems by modelling latent variables that account for the systematic and most Y-relevant 'harmonies' in the X-data.

Thus, with PLS modelling a general tool for relating different types of data tables to each other, is available. This enables the food scientist with one single program to analyze problems ranging from understanding consumer perception of food quality to calibration of advanced multivariate instruments.

ACKNOWLEDGEMENTS

The PLS method could not have been tested out in real life without constructive scientific consultations and skilful technical assistance. Thus, many thanks to M. Rødbotten, S. Hurv, L. Blümlein and E. Risvik (i.e. the sensory group at NINF); H. Russwurm Jr. and M. K. Hårstad (from the chemical group); H. J. Rosenfeld (ensuring agronomical interpretation); P. Lea, C. Irgens and I. Pedersen (from the data group) and L. Bakke and U. Dyrnes (for typing and technical drawings respectively). S. Wold (Umeå University, Sweden) is thanked for continuous support and inspiration.

REFERENCES

Aakesson, C., Persson, T. and von Sydow, E. (1974). Predicting sensory and preference values of aroma and flavour by use of gas chromatographic data. *Proc. IV Int. Congress Food Sci. and Technol.*, Vol. II, pp. 183–9.

Aastveit, A. and Martens, H. (1986). ANOVA interactions interpreted by PLS regression: Straw length in barley genotypes related to climate, *Biometrics*, in press.

Bohlin, L., Egelandsdal, B. and Martens, M. (1985). Relationship between fundamental rheological data and mouthfeel for a model hydro-colloid system. In: *Proc. 3rd Int. Conf. and Industrial Exhibition 'Gums and Stabilisers for the Food Industry'*, Wrexham, UK, July 8–11.

Brown, P. J. (1982). Multivariate calibration, *J. R. Stat. Soc. B*, **44**, 287–308.

Burg, van der E. and Leeuw, de J. (1983). Non-linear canonical correlation. *Brit. J. Math. Stat. Psychol.*, **36**, 54–80.

Canaway, P. R., Piggott, J. R., Sharp, R. and Carey, R. G. (1984). Comparison of sensory and analytical data for cereal distillates. *Proc. Flavour Research of Alcoholic Beverages*, L. Nykänen and P. Lehtonen (Eds), Foundation for Biotechnical and Industrial Fermentation Research, Vol. 3, pp. 301–11.

Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*, Chapman and Hall, London.

Cooley, W. W. and Lohnes, P. R. (1971) *Multivariate Data Analysis*, John Wiley, New York.

EOQC: European Organization for Quality Control (1976). *Glossary of terms used in quality control*, 4th ed.

Esbensen, K. and Wold, S. (1983). SIMCA, MACUP, SELPLS, GDAM, SPACE and UNFOLD: The ways towards regionalized principal components analysis and subconstrained *N*-way decomposition—with geological illustrations. In: *Proc. Nordic Symp. Appl. Stat.*, O. H. Christie (Ed.), Stokkand Forlag Publishers, Stavanger, Norway, pp. 11–36.

Farebrother, R. W. (1984). A note on Fearn's 'Misuse of ridge regression'. *Appl. Stat.*, **33**, 74–5.

Fearn, T. (1983). Misuse of ridge regression in the calibration of a near infrared reflectance instrument. *Appl. Stat.*, **32**, 73–9.

Fjeldsenden, B., Martens, M. and Russwurm, H. Jr. (1981). Sensory quality criteria of carrots, swedes and cauliflower. *Lebensm.-Wiss. u. Technol.*, **14**, 237–41.

Forina, M., Armanino, C., Lanteri, S. and Tiscornia, E. (1983). Classification of olive oils from their fatty acid composition. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 189–214.

Gacula, M. and Singh, J. (1984). *Statistical Methods in Food and Consumer Research*, Academic Press, Orlando, 505 pp.

Geladi, P., MacDougall, D. and Martens, H. (1985). Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.*, May–June, 491–9.

Gower, J. C. (1983). Data analysis: Multivariate or univariate and other difficulties. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 39–67.

Hoffman, D. L. and Young, F. W. (1983). Quantitative analysis of qualitative data: applications in food research. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 69–93.

Jensen, S. Å. and Martens, H. (1983). Multivariate calibration of fluorescence data for quantitative analysis of cereal composition. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 253–70.

Jensen, S. Å., Munck, L. and Martens, H. (1982). The botanical constituents of wheat and wheat milling fractions. I. Quantification by autofluorescence. *Cereal Chem.*, **59**(6), 477–84.

Jöreskog, K. G. and Wold, H. (1982). *Systems under Indirect Observation*, Vols. I and II, North-Holland, Amsterdam.

Kowalski, B. R. (1984). *Chemometrics: Mathematics and Statistics in Chemistry*, D. Reidel, Dordrecht, The Netherlands.

MacFie, H. J. H. (1983). Some useful multivariate techniques relevant to classifying and identifying microbes. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 215–37.

MacFie, H. J. H. and Thomson, D. M. H. (1984). Multidimensional scaling methods. In: *Sensory Analysis of Foods*, J. R. Piggott (Ed.), Elsevier Applied Science Publishers, London, pp. 351–75.

Martens, H. and Jensen, S. Å. (1983). Partial Least Squares regression: A new two-stage NIR calibration method. In: *Progress in Cereal Chemistry and Technology 5a*, J. Holas and J. Kratochvil (Eds), Elsevier Amsterdam, pp. 607–47.

Martens, H. and Næs, T. (1984). Multivariate calibration I. Concepts and distinctions. *Trends in Analytical Chemistry*, **3**, 204–10.

Martens, H. and Næs, T. (1986). Multivariate calibration by data compression. In: *Near Infrared Reflection Spectroscopy*, P. Williams (Ed.), American Association of Cereal Chemists, St. Paul, Minnesota, in press.

Martens, H., Jensen, S. Å. and Geladi, P. (1983a). Multivariate linearity transformation for near infrared reflectance spectrometry. In: *Proc. Nordic Symp. Appl. Stat.*, O. H. Christie (Ed.), Stokkand Forlag Publishers, Stavanger, Norway, pp. 205–33.

Martens, H., Vangen, O. and Sandberg, E. (1983b). Multivariate calibration of an

X-ray computer tomograph by smoothed PLS regression. In: *Proc. Nordic Symp. Appl. Stat.*, O. H. Christie (Ed.), Stokkand Forlag Publishers, Stavanger, Norway, pp. 235–68.

Martens, H., Wold, S. and Martens, M. (1983c). A layman's guide to multivariate data analysis. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 473–92.

Martens, H., Karstang, T., Martens, M., Nilsson, M. B., Hovland, J., Vangen, O., Sandberg, E., Næs, T. and Birth, G. (1986). From dirty data to clean information: multivariate calibration in the analysis of intact samples. *Proc. VIIth Int. Conf. on Computers in Chemical Research and Education*, Garmish-Partenkirchen, June 1985, John Wiley, New York, in press.

Martens, M. (1984). Quality and quality evaluation. *Acta Hort.*, **163**, 15–30.

Martens, M. (1985). Data collection by a microcomputer system in a sensory laboratory, and analysis of sensory data. Lecture notes from *Computers in Sensory Research*, Society of Food Science and Technology, Helsinki, Finland, 10th April.

Martens, M. (1986a). Sensory and chemical quality criteria of stored versus not-stored frozen peas studied by multivariate data analysis. In: *The Shelf Life of Foods and Beverages*, G. Charalamboos (Ed.), Elsevier, Amsterdam, pp. 775–90.

Martens, M. (1986b). Sensory and chemical/physical quality criteria of frozen peas studied by multivariate data analysis. *J. Food Sci.*, **51**(3), 599–603.

Martens, M. and Russwurm, H. Jr. (1980). Evaluation of sensory and chemical quality criteria of dry sausage. In: *Proc. 24th Conf. EOQC*, Warsaw, pp. 88–93.

Martens, M. and Harries, J. (1983). A bibliography of multivariate statistical methods in food science and technology. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 493–518.

Martens, M. and Burg van der, E. (1985). Relating sensory and instrumental data from vegetables using different multivariate techniques. In: *Progress in Flavour Research*, J. Adda (Ed.), Elsevier, Amsterdam, pp. 131–48.

Martens, M., Fjeldsenden, B., Russwurm, H. Jr. and Martens, H. (1983a). Relationships between sensory and chemical quality criteria for carrots studied by multivariate data analysis. In: *Sensory Quality in Foods and Beverages*, A. A. Williams and R. K. Atkin (Eds), Ellis Horwood Ltd, Chichester, pp. 233–46.

Martens, M., Martens. H. and Wold, S. (1983b). Preference of cauliflower related to sensory descriptive variables by partial least squares (PLS) regression. *J. Sci. Food Agric.*, **34**, 715–24.

Martens, M., Risvik, E. and Schutz, H. G. (1983c). Factors influencing preference: A study on black currant juice. In: *Research in Food Science and Nutrition*, Vol. 2, J. V. McLoughlin and B. M. McKenna (Eds), Boole Press, Dublin, pp. 193–4.

Martens, M., Rosenfeld, H. J. and Russwurm, H. Jr. (1985). Predicting sensory quality of carrots from chemical, physical and agronomical variables: a multivariate study. *Acta Agric. Scand.*, **35**(4), 407–20.

Möttönen, K. (1975). *On the Baking Process Dynamics of Wheat*. Part IV,

Technical Research Centre of Finland, Materials and Processing Technology, 11, 40 pp.
Næs, T. (1985a). Comparison of approaches to multivariate linear calibration. *Biometr. J.*, **27**, 265–75.
Næs, T. (1985b). Multivariate calibration when the error covariance matrix is structured. *Technometrics*, Aug.
Næs, T. and Martens, H. (1984). Multivariate calibration II. Chemometric methods. *Trends Analyt. Chem.*, **3**, 266–71.
Næs, T. and Martens, H. (1985). Comparison of prediction methods for multicollinear data. *Commun. Statist.-Simula. Computa.*, **14**(3), 545–76.
Næs, T., Irgens, C. and Martens, H. (1986). Calibration of NIR reflectance instruments. *Appl. Stat.*, in press.
Piggott, J. R. (Ed.) (1984). *Sensory Analysis of Foods*, Elsevier Applied Science Publishers, London.
Piggott, J. R. and Jardine, S. P. (1979). Descriptive sensory analysis of whisky flavour. *J. Inst. Brew.*, **85**, 82–5.
Powers, J. J. (1984). Descriptive methods of analysis. In: *Sensory Analysis of Foods*, J. R. Piggott (Ed.), Elsevier Applied Science Publishers, London, pp. 179–242.
Schiffman, S. S. (1981). Perception of odors of simple pyrazines by young and elderly subjects: a multidimensional analysis. *Pharmacol. Biochem. Behav.*, **14**, 787–98.
Skrede, G., Næs, T. and Martens, M. (1983). Visual color deterioration in blackcurrant syrup predicted by different instrumental variables. *J. Food Sci.*, **48**, 1745–9.
Sundberg, R. (1985). When is the inversed regression estimator MSE superior to the standard regression estimator in multivariate controlled calibration situations? *Statistics and Probabilities Letters*, **3**, 75–9.
Wold, H. (1982). Soft modelling: The basic design and some extensions. In: *Systems under Indirect Observation*, Part II, K. G. Jöreskog and H. Wold (Eds), North-Holland, Amsterdam, pp. 1–54.
Wold, S. (1985). *SIMCA 3B & 3F: Soft Independent Modelling of Class Analogy. Brief program description*, Umeå University, Umeå, Sweden.
Wold, S., Martens, H. and Wold, H. (1983a). The multivariate calibration problem in chemistry solved by the PLS method. In: *Lecture Notes in Mathematics*, no. 973, B. Kågström and A. Ruhe (Eds), Matrix Pencils, Springer-Verlag, Berlin, pp. 286–93.
Wold, S., Albano, C., Dunn III, W. J., Esbensen, K., Hellberg, S., Johansson, E. and Sjöström, M. (1983b). Pattern recognition: Finding and using regularities in multivariate data. In: *Food Research and Data Analysis*, H. Martens and H. Russwurm Jr. (Eds), Applied Science Publishers, London, pp. 147–88.
Wold, S., Albano, C., Dunn III, W. J., Edlund, U., Esbensen, K., Geladi, P., Hellberg, S., Johansson, E., Lindberg, W. and Sjöström, M. (1984). Multivariate data analysis in chemistry. In: *Chemometrics: Mathematics and Statistics in Chemistry*, B. R. Kowalski (Ed.), D. Reidel, Dordrecht, The Netherlands, pp. 17–95.
Wold, S., Wold, H., Dunn III, W. J. and Ruhe, A. (1985). The collinearity problem in linear and non-linear regression. The partial least squares (PLS) approach to generalized inverses, in press.
Wold, S., Martens, H. and Martens, M. (1986). Finding the main common information in multi-table sensory data by Consensus Partial Least Squares (CPLS) modelling. In preparation.
Young, F. W. (1981). Quantitative analysis of qualitative data. *Psychometrika*, **46**, 357–88.
Young, F. W. (1985). Alternating Least Squares. Building individuals into the model. Lecture at Int. meeting on Relating sensory data to physical and chemical measurements, Society of Chemical Industry & Royal Statistical Society, London, 29th April.
Young, F. W. and Lewyckyj, R. (1979). *ALSCAL-4 User's Guide*, 2nd edn, Data Analysis and Theory Association, Carrboro, N. C.